

PROCEEDINGS

OF THE

ROYAL SOCIETY OF LONDON

SERIES A

CONTAINING PAPERS OF A MATHEMATICAL AND
PHYSICAL CHARACTER.

VOL. CXXXVIII.

LONDON:

PRINTED FOR THE ROYAL SOCIETY AND SOLD BY
HARRISON AND SONS, LTD., ST. MARTIN'S LANE,
PRINTERS IN ORDINARY TO HIS MAJESTY.

DECEMBER, 1932.

LONDON:

HARRISON AND SONS, LTD., PRINTERS IN ORDINARY TO HIS MAJESTY,
ST. MARTIN'S LANE.

CONTENTS.

SERIES A. VOL. CXXXVIII.

No. A834—October 1, 1932.

PAGE

The Modes of Vibration of Quartz Piezo-Electric Plates as Revealed by an Interferometer. By The Late W. D. Dye, F.R.S. Communicated by Sir Joseph Petavel, F.R.S. (Plates 1-3).....	1
Theory of Electric Charge. By Sir Arthur Eddington, F.R.S.....	17
The Viscosity of a Fluid Containing Small Drops of Another Fluid. By G. I. Taylor, F.R.S.	41
On the Theory of Errors and Least Squares. By H. Jeffreys, F.R.S.....	48
Contributions to the Mathematical Theory of Epidemics. II.—The Problem of Endemicity. By W. O. Kermack and A. G. McKendrick. Communicated by Sir Gilbert Walker, F.R.S.	55
The Absorption Spectrum of Nitrous-Oxide and the Heat of Dissociation of Nitrogen. By A. K. Dutta. Communicated by M. N. Saha, F.R.S.....	84
The Mechanism of the Initiation and Propagation of Detonation in Solid Explosives. By W. Taylor and A. Weale. Communicated by F. G. Donnan, F.R.S.....	92
Mobility of Alkali Ions in Gases. By C. F. Powell and L. Brata. Communicated by A. P. Chattock, F.R.S.....	117
The Distribution of Suspended Particles Under Gravity. By C. M. McDowell and F. L. Usher. Communicated by R. Whytlaw Gray, F.R.S. (Plate 4).....	133
Interferometric Measurements in the Spectrum of Krypton. By C. V. Jackson. Communicated by A. Fowler, F.R.S.....	147
Dipole Moments and Molecular Structure. Part III.—The Oxychlorides of Sulphur. By J. W. Smith. Communicated by F. G. Donnan, F.R.S.	154
The Swelling of Charcoal. Part II.—Some Factors Controlling the Expansion Caused by Water, Benzene and Pyridine Vapours. By D. H. Bangham, N. Fakhoury, and A. F. Mohamed. Communicated by D. L. Chapman, F.R.S.	162
Theory of Uncoupling and Formulæ for the Stark Effect in H_2 . By J. K. L. MacDonald. Communicated by R. H. Fowler, F.R.S.....	183
The Polarity of Thunderclouds. By E. C. Halliday. Communicated by C. T. R. Wilson, F.R.S. (Plates 5-10).....	205
Inelastic Electron Scattering in Gases.—I. By C. B. O. Mohr and F. H. Nicoll. Communicated by Lord Rutherford, F.R.S.....	229
Further Experiments on Superconductivity with Alternating Currents of High Frequency. By J. C. McLennan, F.R.S., A. C. Burton, A. Pitt and J. O. Wilhelm	245

No. A835—November 1, 1932.

PAGE

Adsorption. A Study of Availability and Accessibility. By Sir William B. Hardy, F.R.S., and M. Nottage.....	259
On Plasticity and Creep in Solids. By H. Jeffreys, F.R.S.....	283
The Explosive Oxidation of Carbon Monoxide at Lower Pressures. By G. Hadman, H. W. Thompson and C. N. Hinshelwood, F.R.S.....	297
The Lower Pressure Limit in the Chain Reaction between Hydrogen and Oxygen. By C. N. Hinshelwood, F.R.S., and E. A. Moelwyn-Hughes.....	311
The γ -Rays of Thorium B and of the Thorium C Bodies. By C. D. Ellis, F.R.S. (Plate 12)	318
The Theory of Wave Resistance. By T. H. Havelock, F.R.S.....	339
The Resistivity of Polycrystalline Wires in Relation to Plastic Deformation, and the Mechanism of Plastic Flow. By E. N. da C. Andrade and B. Chalmers. Communicated by L. N. G. Filon, F.R.S.....	348
The Photochemistry of Phosphine. By H. W. Melville. Communicated by J. Kendall, F.R.S.....	374
Bands due to the Hydrogen Molecule: The $2p^3\Pi$ Bands of Hydrogen. By I. Sandeman. Communicated by H. S. Allen, F.R.S.....	395
The Structure of Surface Films. Part XVI.—Surface Potential Measurements on Fatty Acids on Dilute Hydrochloric Acid. By N. K. Adam and J. B. Harding. Communicated by F. G. Donnan, F.R.S.....	411
On the Surface Potentials of Unimolecular Films. Part IV.—The Effect of the Underlying Solution and Transition Phenomena in the Film. By J. H. Schulman and A. H. Hughes. Communicated by E. K. Rideal, F.R.S.....	430
The Behaviour of Electrolytes in Mixed Solvents. Part IV.—The Free Energy of Zinc Chloride in Water-Alcohol Solutions. By R. T. Hamilton and J. A. V. Butler. Communicated by J. Kendall, F.R.S.	450
The Passage of Neutrons through Matter. By H. S. W. Massey. Communicated by R. H. Fowler, F.R.S.	460
The Large Angle Scattering of Electrons in Gases.—II. By C. B. O. Mohr and F. H. Nicoll. Communicated by Lord Rutherford, F.R.S.	469

No. A836.—December 1, 1932.

The Physical Principles of the Quantum Theory. By G. Temple. Communicated by S. Chapman, F.R.S.	479
A Unique Electrode Potential Characteristic of a Metal, and a Theory for the Mechanism of Electrode Potential. By A. L. McAulay and E. C. R. Spooner. Communicated by T. M. Lowry, F.R.S.....	494
Triboelectricity and Friction. VII.—Quantitative Results for Metals and other Solid Elements, with Silica. By P. E. Shaw and E. W. L. Leavey. Communicated by Sir William Hardy, F.R.S.	502

	PAGE
An X-Ray Investigation of Normal Paraffins near their Melting Point. By A. Müller. Communicated by Sir William Bragg, F.R.S. (Plate 13).....	514
Investigations in the Infra-Red Region of the Spectrum. Part VII.—An Infra-Red Grating Spectrometer as a Double Monochromator. By A. B. D. Cassie and C. R. Bailey. Communicated by F. G. Donnan, F.R.S.....	531
The Resonance Spectrum of Hydrogen. By K. R. Rao and J. S. Badami. Com- municated by A. Fowler, F.R.S. (Plate 14).....	540
The Calculation of the Terms of the Optical Spectrum of an Atom with one Series Electron. By J. McDougall. Communicated by R. H. Fowler, F.R.S.....	550
Rotational Uncoupling, with application to the Singlet Hydrogen Bands. By P. M. Davidson. Communicated by O. W. Richardson, F.R.S.	580
The Theory of Metals. I. By A. H. Wilson. Communicated by R. H. Fowler, F.R.S.	594
The Torsion and Flexure of Shafting with Keyways or Cracks. By W. M. Shepherd. Communicated by L. N. G. Filon, F.R.S.	607
The Oxidation of Sulphur at Low Pressures. By A. Ritchie and E. B. Ludlam. Communicated by J. Kendall, F.R.S.	635
The Internal Conversion Coefficient for Radium C. By H. R. Hulme. Communi- cated by R. H. Fowler, F.R.S.	643
A Theory of the Internal Conversion of γ -Rays. By H. M. Taylor and N. F. Mott. Communicated by R. H. Fowler, F.R.S.	665
The Rate of Burning of Colloidal Propellants. By F. R. W. Hunt and G. L. Hinds. Communicated by H. C. Plummer, F.R.S.	696
The Relationship between Viscosity, Elasticity and Plastic Strength of Soft Materials as Illustrated by some Mechanical Properties of Flour Doughs. I. By R. K. Schofield and G. W. Scott Blair. Communicated by Sir John Russell, F.R.S. (Plate 15)	707
—————	
Index	719



PROCEEDINGS OF THE ROYAL SOCIETY.

SECTION A.—MATHEMATICAL AND PHYSICAL SCIENCES.

The Modes of Vibration of Quartz Piezo-Electric Plates as Revealed by an Interferometer.

By the late W. D. DYE, D.Sc., F.C.G.I., F.R.S., The National Physical
Laboratory.

Prepared for publication by P. VIGOUREUX, M.Sc.

(Communicated by Sir Joseph Petavel, F.R.S.—Received June 6, 1932.)

[PLATES 1-3.]

I. *Introductory.*

During the past five years considerable development has taken place in the piezo-electric quartz oscillator and resonator, chiefly in connection with their application to the production and measurement of radio-frequencies of great constancy or precision.

A considerable literature has grown up dealing with the circuit arrangements and other practical details of construction found desirable or necessary by various workers in order to secure a satisfactory performance in the assemblage constituting a complete resonator or oscillator. This information is scattered throughout a considerable number of publications and for the most part must be considered as being of a disjointed nature. Bibliographies of the subject have, however, been drawn up by Cady, and more recently a critical resumé of the important literature has been collected in a monograph published by H.M. Stationery Office.*

Some of the investigations have been of a more fundamental nature, directed towards a better understanding of the nature of the structure of quartz and of the vibrations which occur in rods and plates. The names of Bragg and Gibbs,

* P. Vigoureux, "Quartz Resonators and Oscillators." H.M. Stationery Office, London (1931).

Meissner, Giebe and others stand out in this connection. Other investigators have concentrated on the theoretical and practical aspects of the subject, amongst whom may be mentioned Cady, Van Dyke, Vigoureux, Terry, and Dye, the work of the last author being mainly an experimental verification of some electrical equivalent constants implicit in the theory of a quartz resonator.

Certain other workers have sought to elucidate the nature of the vibrations in quartz plates by means of glow discharges, sand figures, air currents, etc., displayed by the vibrating surfaces of the cut crystal. Others again have made use of the temperature coefficients of frequency, and of the discontinuities of frequency with temperature, as a means of deducing the laws of vibration of discs and plates cut in special relation to the crystal axes. A very valuable and interesting paper by Lack throws much light on the whole subject and should be read by all interested in it.

A study of this mass of information shows that a good deal is known about the behaviour of resonators and oscillators, and that in the case of thin rods the information may be considered to be nearly complete and in accordance with theoretical predictions. In the case of plates, however, the position is by no means so satisfactory; theory and experiment are very discordant, and the vagaries of oscillators with respect to reproducibility and permanence are still perplexing. The very recent work on torsional oscillations by Hund and Wright shows that much of an obscure nature occurs and is not understood.

The present paper has for its object the investigation of the actual type of motion which occurs in plates of various shapes when vibrating at any of their numerous natural free resonant frequencies.

II. *Statement of the Problem.*

The piezo-electric properties of quartz are sufficiently well known not to need detailed description. It will suffice to state that cut crystal used in quartz oscillators and resonators is in general of two forms:—

- (a) The rod form, with length perpendicular to the electric and optic axes. The depth is usually small compared with the length, and is parallel to the optic axis. The thickness is usually the smallest dimension and is parallel with an electric axis. The electrodes apply the field in the direction of the electric axes.
- (b) The plate form results when the dimension called the depth in (a) is made comparable to the length. In such cases the boundary of the surface of the plate may assume various forms, of which the circle and rectangle

are commonest. The optic axis nearly always lies parallel to the surface, but the thickness direction of the plate is not necessarily parallel to an electric axis. In a common alternative mode of cutting the plate the optic axes and an electric axis lie in the plane of the plate.

Other forms of crystal which have been used are rectangular blocks and thick toroidal rings in which all dimensions are comparable. The latter have been used by Morrison, and give a very small temperature coefficient when cut so that the axis of the toroid is perpendicular to both the optic and an electric axis. The small temperature coefficient of such ring-shaped oscillators is said to be due to a compensatory effect produced by the shear component of the vibration.

As mentioned above the mode of vibration of a rod, when longitudinal, is of simple form and may be said to be fairly well understood. Even in this case, however, there must be a shear component, owing to the differences in the elastic modulus in various directions perpendicular to the electric axis. There is room for further investigation into the effects of this shear component on the temperature coefficients of rods of various ratios of length to depth.

When plates are used, the mode of vibration desired is usually that which ideally consists of a pure compressional vibration of gravest mode in the thickness direction. In such an ideal case there would be a median plane as node, and the two flat surfaces of the plate would approach and recede from each other with pure translatory motion. Many experiments have shown that the motion is by no means of this simple kind, and a number of guesses have been made from time to time as to what the motion really is like.

It occurred to the author that a hopeful line of experiment would be to use the principle of interference as an aid to the examination of the mode of deformation of the surface of a vibrating plate. Some preliminary experiments showed that the method possessed considerable possibilities, and a simple form of interferometer was accordingly constructed, as described below.

III. Description of Interferometer.

The principle of the method of examination by means of an interferometer is to prepare the piezo-electric plate of any desired shape and size and to grind and polish one of the flat surfaces to a true optical plane. This surface is then brought into close approximation to a reference plane-polished surface in order that interference bands may be produced and observed. The other essential requirement is an electrode system with means of applying to the

plate an alternating electric field smoothly adjustable in frequency. The arrangement of the apparatus is seen in fig. 1.

The quartz plate *Q* rests freely upon a brass disc *D* ground to an approximate plane, used as lower electrode. Through the base-plate of the instrument pass three adjustable screws which support on their steel ball-ends a thick glass plate *P*. The under surface of *P* is polished flat, and constitutes the reference surface. The upper surface of *P* has purposely been given a small inclination to the lower, so that any light reflected from it cannot interfere with light reflected from the reference surface.

A thin square brass frame is held by soft wax to the middle portion of the lower surface of *P*. It supports a grid of fine silver wire, wound to and fro across one face of the frame and round small pins let in on two opposite sides. The diameter of the wire is about 0.02 mm. and the pitch of the winding about 1 mm. This grid lies flat against the reference surface of *P*, and thus forms an upper electrode which only slightly obstructs the light. In early experiments a wire gauze was made very fine by dissolving almost the whole of it in acid. It yielded satisfactory results, but obstructed more light than the present parallel grid, by which it is replaced in the later work.

The rest of the apparatus consists of the arrangements for projecting light normally on the two interfering surfaces, and for receiving part of the reflected light into the eye. The light of a mercury vapour lamp at *S*, after passing through a filter, if desired, is concentrated into a pin-hole at *H*, and passes down through a piece of ordinary glass plate *G*, on to a lens *L* of such focal length that the emergent light is parallel. After reflection from the two interfering surfaces the light passes back along its original path to the glass plate *G*, where part of the convergent beam is reflected and brought to a focus just outside the pin-hole *O*. The illuminated surface can be seen at an effective distance of about 25 cm. by looking through *O*. Photographs can be taken by replacing the eye at *O* by the bellows and plate carrier of a camera. As explained later the mercury vapour lamp can be replaced by a helium tube as the source of illumination, in which case the capillary of the helium tube is placed immediately above the pin-hole *H*.

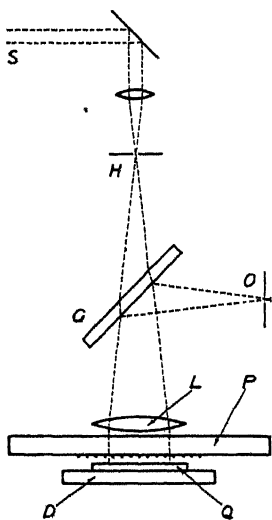


FIG. 1.

In practice an air gap of about 0.2 mm. is used between the grid electrode and the upper surface of the quartz plate under examination. By adjustment of the screws supporting P the angle between the surfaces is reduced until interference fringes are seen. They can be given any desired pitch and direction by suitable manipulation of the screws. If both surfaces are flat, the fringes will be straight, parallel and equi-spaced.

Quartz plates of a few centimetres diameter and of thickness greater than 2 mm. can be ground and polished to a flatness of one-tenth of a wave-length of green light without difficulty. By very careful adjustment of the reference plate a black band can be obtained covering the whole area of the specimen. For many purposes the bands may be pitched about 1.5 mm. apart. In this condition the surface of the specimen need not be of such perfection of flatness as indicated above.

With bands covering the whole surface of the quartz plate, it will readily be understood that statical deformation from flatness of the whole or any part of the surface will cause a deflection of the bands in the corresponding region. Thus for a spherical shape the bands will become circular about a common centre. A cylindrical deformation will give bands straight but of unequal spacing, if the axis of the cylinder is parallel to them. If the axis of the cylinder is perpendicular to the original direction of the bands, they will become parts of ellipses, but will remain equi-spaced. If any part of the surface is displaced normally to itself the bands in this region will become displaced perpendicularly to their direction, which will remain unchanged. Similarly a change of slope of the surface at any part will produce a corresponding slewing round of the bands or change in their distance apart. It will thus be seen that the conformation of the bands is a direct measure of the form of the surface, although in some cases this form cannot be readily visualised by inspection of the bands.

If the quartz plate is set vibrating piezo-electrically in one of its many modes, the surface will vibrate in various ways, and in general certain regions will have a component of vibration in the direction perpendicular to the plane. A rapid oscillatory movement of the parts of the bands in these regions will occur, and they will disappear from view entirely. Regions which remain at rest will, on the other hand, be revealed by the unchanged appearance of the bands. The whole surface will therefore be mapped out into nodal and anti-nodal regions well differentiated from one another. The method bears a close analogy to that of Chladni in which sand figures are produced on vibrating plates; the present method is, however, more sensitive and more precise.

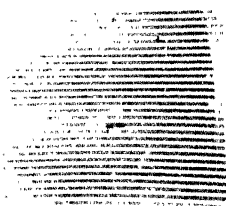
The appearances described above refer to cases in which the source of light is continuous, *e.g.*, a mercury vapour lamp. It is possible, however, as described later, to use a source of light intermittent at the same frequency as the vibration of the plate; when such a source of illumination is used a stroboscopic effect is obtained, and instead of the bands being broken up into visible and invisible portions, they are seen in their entirety, since the illumination occurs for a single very short fraction of each cycle of the vibration of the plate. If the instant of illumination is that corresponding to maximum deformation of the plate, the bands will be seen at their maximum displacement. The whole surface will be seen covered with curved and displaced bands, which are virtually a contour map of the given instant. Such intermittent illumination can be obtained by means of a helium tube with a capillary portion which can be supported directly over the pin-hole H.

Each of the two methods of observation has its advantages. Thus for observations on the exact form and location of the nodes and antinodes, continuous illumination is preferable. But to see the actual deformation and to measure its amount in various parts of the surface, intermittent illumination is much to be preferred.

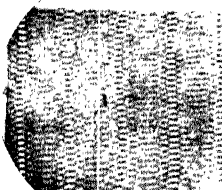
The apparatus used to excite the plate and to illuminate it intermittently consists of a valve source of electrical oscillations coupled to the interferometer and to the helium tube. The source was of conventional type employing a small transmitter valve operating on a 400-volt anode supply. The condensers of the oscillatory circuit were variable air condensers and fixed mica condensers. The frequency could be varied smoothly between the limits 7 kc. and 3000 kc. per second and could, if necessary, be adjusted to a precision of one part in a million.

The two electrodes of the quartz system were usually connected in parallel with the condenser of the oscillatory circuit, but in some cases, in order to avoid too great a voltage, they were connected to an inductance coil brought near to the main inductance in the oscillatory circuit.

The arrangements for illuminating the helium tube intermittently, at the source frequency, consist of a separate tunable circuit of relatively large inductance and small capacity, coupled inductively to the main source. The A.C. voltage thus obtained has a value of about 400 volts. A D.C. voltage for polarising purposes is connected in series with the A.C. supply, and the helium tube is connected through resistances of about 10,000 ohms on each side to the resultant pulsating supply. By this means the tube gives one flash each cycle, and the phase of this with respect to the vibrating surface of the quartz plate can be adjusted by slightly detuning the oscillatory circuit.



(a)



(b)



(c)

FIG. 2.



(a)

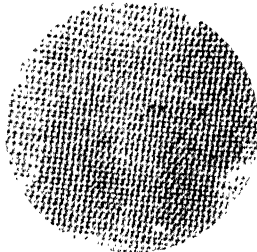


(b)



(c)

FIG. 3.



(a)



(b)



(c)

FIG. 4.

The instant of illumination can thus be adjusted to that at which the quartz plate is instantaneously at rest at its maximum deformation. This instant is particularly favourable for the two reasons that (a) the deformation is a maximum, and (b) the finite duration of the illumination causes the least blurring, since the change in deformation which occurs during this short time is least. The example shown in fig. 2a, b and c, Plate 1, serves to illustrate the points just described. Fig. 2a shows the appearance of the bands for the case of a rectangular plate of dimensions $4.4 \times 3.6 \times 0.7$ cm. at rest, the bands being adjusted to parallelism with the longer edge. In fig. 2b is shown the appearance under continuous illumination from the mercury arc, using the violet lines of the spectrum. The upper electrode was a thin phosphor bronze strip, seen as a black line about midway between the second and third nodal lines. It will be seen that the bands are obliterated in all but five narrow segments of each, which together form the five nodal lines proper to the mode of vibration. The mode of vibration is clearly revealed as one of practically pure flexure of overtone type.

A closer examination of the figure will show that all trace of bands is obliterated in long narrow belts on each side of the nodes. These are regions in which the amplitude of vibration is such that the to-and-fro vibration of the bands is just the amount necessary to give an average illumination constant for any point within the region. Adjacent to these belts are regions, also in the form of parallel belts, where shadowy bands are seen. It will be observed that these bands are displaced by half a band width from the true nodal bands. Beyond this again are belts where the bands are quite invisible; and then, in the middle of the vibrating segments, there are again shadowy bands. It will be understood that the appearance of the shadowy bands occur at certain amplitudes which are dependent only upon the wave-length of the light used. It will thus be seen that the interference method, even with continuous illumination, can be utilised to give quantitative information regarding the form of deformation in simple cases.*

By a knowledge of the intensity distribution of light in the band system and from an assumption of a sine-wave law of band displacement when vibrating, it should be possible to calculate the displacement corresponding to uniform illumination along the belts just described. There is, however, considerable uncertainty about the light intensity law, and the expression can therefore only be conveniently solved graphically. Fortunately the method of inter-

* Cf. Osterberg, 'Proc. Nat. Acad. Sci. Wash.,' vol. 15, p. 892 (1929).

mittent illumination overcomes all difficulties regarding the actual form of the displacement, and as shown in fig. 2c, taken with intermittent light from a helium tube, the contour of the surface can be directly seen, and measured if desired.

It will be observed that in figs. 2b and 2c the bands have not perfect uniformity, but are more sharply defined in certain belts than in others. In fig. 2b this effect is due to the presence of the blue and the two violet lines of the mercury arc, all of which are very actinic. The systems of bands from these colours reinforce one another in certain regions and appear individually in others. Filters to separate these lines are not easily obtainable, and the green line, which can of course be separated, requires much longer exposure or special photographic plates. The effect is not of serious consequence, however, so that, except in a few later cases, the photographs taken with the arc all show the effect. In the case of the helium tube illumination used for the stroboscopic photographs this effect is much less pronounced, because, under the conditions of use, one line in the violet is much more active than the other.

IV. *Experimental.*

Having set up the interferometer and the source arrangements for providing the smoothly variable high-frequency voltage, the natural procedure was to cut a few quartz plates and explore the responses obtained when the voltage was applied and its frequency smoothly varied. A number of circular and rectangular plates were therefore prepared from selected quartz. The block was first tested optically for freedom from optical twinning. Plates were then cut from it, and in some cases were tested piezo-electrically by a simple apparatus with which a small electrode was pressed with constant force on to various parts of the plate. A string electrometer measured the charge produced, and so enabled an exploration of the area of the plate before the disc or rectangle was sent from it. The discs and rectangular plates were cut and ground to a high degree of parallelism, and the edges were ground true to the surface. In most cases only one face of the plate was polished, but in some cases for special reasons both faces were polished. Considerable care was taken in the polishing to work the surface plane to at least one-half wave-length of green light, and in some cases still greater care was taken to obtain flatness to one-tenth wave-length.

The first plate tried was a disc of about 4 cm. diameter and 7 mm. thick. It was set up in the interferometer as described and the electrodes were con-

nected up to the source. As soon as the source was slowly and smoothly varied in frequency it was seen that a large number of response vibrations were excited in the disc. The number of such responses is very great indeed, and in traversing a range of frequencies from 2×10^6 down to 5×10^4 per second there were literally hundreds of modes of vibration in the disc under experiment.

The modes of vibration of such a disc, as well as the modes of rectangular plates, form well-defined groups which are clearly distinguishable by the interference patterns formed. They may be classified as follows :—

- (1) Compressional vibrations in the direction perpendicular to the plane of the plate. These vibrations predominate at frequencies above that at which the half wave-length equals the thickness of the plate.
- (2) Compressional vibrations in various directions parallel with the plane of the plate.
- (3) Flexural vibrations about axes lying in various directions in the plane of the plate.
- (4) Torsional vibrations of pure and of complex kinds.
- (5) Patterned flexural and torsional mixed vibrations.
- (6) Vibrations of types more complicated than the above.

Each of these kinds of vibration may be of fundamental or of overtone type, but as a general rule it may be stated that when an overtone type of one of the kinds of vibration falls within the region of frequency of the fundamental of another kind, the former is no longer maintained in pure form.

Thus, for example, the overtone flexural vibration with 5 nodes shown in fig. 2*b* has a frequency considerably lower than that of the fundamental compressional kind under heading (1) above. Successive overtone modes with 6, 7 and 8 nodal lines can also be freely excited in the plate of fig. 2*b*, but above 8 nodes the frequency is in the region corresponding to vibrations of type (1) and the pure overtone flexural vibration with 9 nodes cannot be elicited.

Owing to the extensive new field opened up by this method of examination, it was thought desirable simply to try various shapes and sizes of plate in a tentative fashion, so as to make a rough survey of the possible modes of vibration, before carrying out a more careful and detailed investigation along some particular direction. The following figures and remarks refer to a number of plates prepared one after another, and serve to show the variety and beauty of the modes obtained in this somewhat casual manner.

Figs. 3*a* and 3*b*, Plate 1, show a disc vibrating in its most important mode, compressionally, at a frequency corresponding to its thickness. This disc was 7.5 mm. thick and 48 mm. in diameter, and its plane was perpendicular to an electric axis. Fig. 3*b* taken with intermittent illumination, shows in a striking manner the crispation of the surface due to the mixed longitudinal compressional component (in direction perpendicular to the plane) and transverse compressional components in two directions in the plane of the disc. There is a considerable regularity in the undulations of the surface, which could not be easily inferred from fig. 3*a*, taken with continuous illumination. The undulations shown in fig. 3*b* are two intersecting sets, one set is disposed along five parallel lines running upwards from left to right at about 30° to the horizontal and nearly parallel with the interference bands. The second set of corrugations runs downwards from left to right at about 30° to the horizontal. In this direction the corrugations have their maximum amplitude. This can be seen by reference to fig. 3*a*, which may now be considered briefly. The nodes revealed in this photograph are seen to be somewhat irregular. There is a diametral belt strongly vibrating as a whole but not of uniform amplitude, whilst on either side are nodes of looped or of figure-of-eight formation. It was not certain at this stage whether the transverse compressional vibrations were actually compressional or not, and accordingly another disc was prepared and polished on both sides. This disc was supported upon a polished plate with small pieces of very thin foil at two places near the rim, the two surfaces being in contact at a point opposite the foil. A very thin wedge-shaped air space was thus left between the plates: this allowed a set of interference fringes corresponding to the lower surface of the quartz plate to be formed.

The upper glass reference plate was adjusted to give interference bands with the upper surface of the quartz plate. These bands were adjusted to be of about the same pitch and at right angles to the bands from the lower surface. The two intersecting bands are shown for the condition of rest in fig. 4*a*, Plate 1. By this means it is possible to obtain information regarding the motions of the two surfaces at the same instant.

Fig. 4*b* shows the appearance under continuous illumination when the disc is vibrating in its strongest mode similarly to the plate of fig. 3*a*. The nodal and anti-nodal areas are well seen, and are again found to be of circular or ring-shaped form. It will be seen that in every part of the anti-nodal regions both sets of fringes are visible. This proves conclusively that the nodal regions on the upper surface and those on the lower surface are accurately opposite one another. The photograph with intermittent light shown in fig. 4*c* was

then taken, with some difficulty. From the directions of the deformations of an interference fringe of the upper surface at any chosen point, it can be stated whether this point at the instant of illumination lies above or below the rest position, the accessory information required being, of course, the directions in which the air wedge is becoming thicker. At this same selected region of the disc, but on the opposite face, the direction of deformation of the corresponding interference fringe at this point can be noted. It can thus be determined whether, at any instant, at any two points opposite each other on the disc, the surfaces are approaching each other or receding from each other, or are moving in the same direction. If they are approaching or receding, the motion must be one of compression, and if they are moving together it must be one of flexure. In the actual case shown the motion was proved to be compressional. Further experience with plates and discs showed that whenever the vibration has the fundamental compressional mode in the thickness direction as the chief component, the accompanying transverse overtone components are also compressional. In such cases the nodes are usually closed loops.

In fig. 3c, Plate 1, is shown an accessory mode of longitudinal compressional vibration of the disc as shown in figs. 3a and 3b. The frequency differs by only a few parts in a thousand from that shown in fig. 3a. The mode is a very different one, and is of smaller general amplitude in the vibrating segments. There is considerable symmetry, and the looped form of node is very well shown.

In fig. 5a, Plate 2, is shown under continuous illumination a disc much thinner than that of fig. 3a. It has a ratio of diameter to thickness of 11 : 1, and shows therefore a closer type of nodal pattern, of a more or less honeycomb form, but not sufficiently regular to enable a precise definition of it to be made. The closer formation of the nodes and loops is to be expected in view of the higher frequency of this thinner disc. Fig. 5b shows the same disc, again vibrating in its principal longitudinal mode, but the bounding curved surface, which in fig. 5a was cylindrical, has been changed by putting a considerable chamfer on each edge, so that the curved surface is virtually composed of a central cylindrical part with a conical part on each side. The change in the mode of vibration is most marked. Instead of a more or less uniform honeycombed nodal pattern, there is now a circumferential nodal pattern with a central portion all of which is vibrating. This effect is general with discs, and results in an enhanced piezo-electric reaction upon the electrode system, since the longitudinal component represents a larger proportion of the total mixed vibration than in the case of fig. 5.

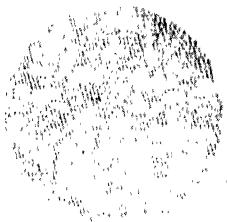
In fig. 6, Plate 2, is shown a disc considerably thicker than those of figs. 3 and 5. The ratio of diameter to thickness is about 4.5 : 1, thus giving a more open nodal pattern. The mode of vibration is, however, not unlike the disc of fig. 3a, in that there is a complete diametral vibrating segment with looped nodal regions on either side.

The next figure, fig. 7, Plate 2, shows the behaviour, with intermittent light, of a rectangular plate, about 5 cm. long by 4 cm. wide, in the optic axis direction, and about 1 cm. thick. The vibration is again the strongest fundamental longitudinal one in the thickness direction. This example has been chosen to illustrate the difficulty of visualising the exact nature of the deformation of the plate, and is referred to later in the Appendix. It will suffice here to state that the vibration shows an unsuspected symmetry, and is of a type which is curious and difficult to explain.

Fig. 8 shows a square plate about 6 mm. thick in its strongest compressional mode of vibration in the thickness direction. The direction of the interference bands has been chosen parallel to an edge in order to bring out the partial regularity of the transverse compressional overtone modes about the two axes nearly parallel to any adjacent edges. The surface is thus divided up into hills and valleys, but these are not uniform, since the whole of the upper half of the plate is vibrating except for a small node in the left-hand corner.

We now turn to some quite different modes of vibration of discs and rectangular plates, at frequencies lower than those corresponding to compressional vibrations in the thickness direction. Figs. 9a and 9b, Plate 2, for example, show the disc already represented in fig. 3, this time vibrating in a segmented overtone flexural mode. The two modes occur quite near to one another in frequency, and it will be seen that the peripheral segments of the one mode lie in the regions between those of the other mode. The two modes are complementary to one another, and probably represent two coupled systems of equal frequency and small coefficient of coupling.

The symmetry of this and similar modes of vibration is practically perfect, and is a feature of flexural vibrations generally. The segments and central closed portions are in phase with one another, and are in opposite phase to the remainder of the surface. The rim thus presents a crinkled appearance. This is well seen in the model carved in relief which is referred to in the Appendix. Theoretically, with perfect uniformity of the electric field and of the quartz, it should not be possible to elicit vibrations of the type here shown, since there should be no piezo-electric reaction for a flexural mode. Actually, however, there is always a small non-uniformity which allows a resultant



(a)

FIG. 5.



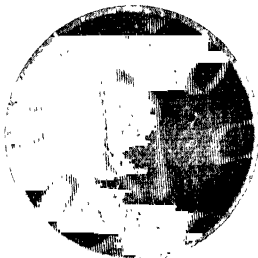
(b)



FIG. 6.



FIG. 7.



(a)

FIG. 9.



(b)



FIG. 8.



(a)

FIG. 10.

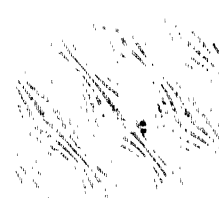


(b)



(a)

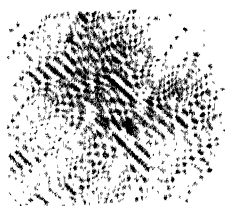
FIG. 11.



(b)

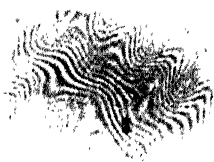


FIG. 12.



(a)

FIG. 13.



(b)

FIG. 14.



FIG. 15.



FIG. 16.



FIG. 17.



FIG. 18.

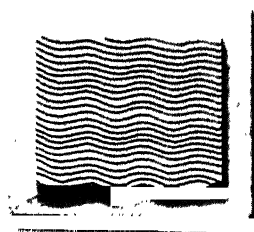


FIG. 19.



FIG. 20.

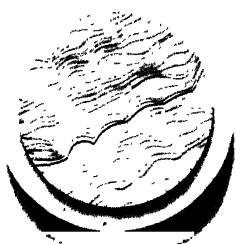


FIG. 21.

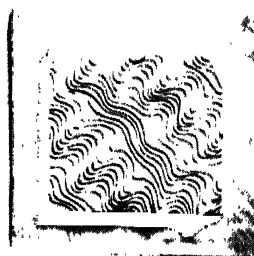


FIG. 22.

maintaining force, and with suitably disposed electrodes, arranged to produce electric fields favourable at all parts of the disc to such a mode, it can be excited very strongly.

Another, simpler, flexural mode of vibration, which can easily be excited in all discs, is that in which the nodes are radial and always even in number. An example is not given as it is a very obvious type.

In figs. 10*a* and 10*b*, Plate 2, are shown photographs, in continuous and intermittent light respectively, of the disc of fig. 3 vibrating in a mode with annular nodes combined with one diametral one. The conformation of the surface can be well judged by examination of fig. 10*b*, which shows that the amplitude of the central two segments is much greater than that of any other part of the disc. In this and in the preceding example the disc remains sensibly of constant thickness throughout the cycle of vibration.

The lower frequency modes of vibration of rectangular plates take many remarkable forms; thus figs. 11*a* and 11*b*, Plate 2, show a rectangular plate in which the nodal lines form two well-defined groups of parallel belts dividing the surface into lozenge-shaped vibrating segments each in opposite phase to its neighbour; the surface is thus divided into mounds and hollows in very regular formation. The shorter side of the plate is parallel to the optic axis. The equal spacing of the two sets of nodal lines indicates that their directions are so related to the optic axis that the Young's modulus along them is equal.

Another type of vibration, which can always be obtained in plates of somewhat elongated form, is that shown in fig. 12. This is an overtone torsional vibration. There is a central backbone and a set of fins radiating from it. A line parallel with the shorter edge, if advanced perpendicular to itself in the plane of the surface, would experience a tilting up and down as it passed through the anti-nodal regions. The upper electrode in this case was the single strip shown as a grey rectangle in the central lower segment. Powerful vibrations of this type can be obtained by using suitably disposed circular electrodes in the appropriate segments.

An entirely different type of vibration found in rectangular plates is shown in figs. 13*a* and 13*b*, Plate 3. The main nodal lines are here closely parallel and very sharply defined. The vibrations in the anti-nodal spaces between them are very intense, and frequently cause fractures of the plate. This type of vibration is akin to an overtone torsional type with axis of torsion running from the top left-hand to the bottom right-hand corner. Although there is no actual nodal line in this direction, it can be seen on reference to the fig. 13*b*, taken by intermittent light, that there is a diagonal region without much

vibration. In fig. 14, Plate 3, the same plate is shown vibrating at a higher frequency in a still more complicated fashion. The system consists of three backbone nodes running across at an angle of about 70° to the horizontal. These are joined laterally by numerous rib-like nodes all roughly parallel to one another, giving a compound overtone torsional quality to the vibration. This description is, however, by no means perfect, and there are other ways of regarding the motion. The distance separating the sharply defined nodal lines is about equal to the thickness of the plate.

In conclusion, further types of vibration are shown in figs. 15-18, Plate 3. It will be noted that in these the nodal and anti-nodal regions are distinguished by light and shade only, and that no interference bands are seen. This is accomplished by adjusting the reference interfering surface so that it is strictly parallel to the upper polished surface of the quartz plate. It can thus be arranged that a black band fills the entire area when the plate is not vibrating. For this adjustment highly accurate flat surfaces are necessary, together with considerable care and steadiness of conditions during exposure. The ideal has not been quite realised in fig. 15, Plate 3, which shows a black band occupying the right half of the photograph and a bright band the left half; there is, however, not much loss of effect due to this. The other three pictures are nearly perfect in this respect.

Fig. 15 shows a type of vibration in which the anti-nodes are arranged round an ellipse. There is a circumferential nodal belt and a series of conjugate nodes roughly radially disposed. In fig. 16, Plate 3, the vibration is approximately conjugate in type to that of fig. 15. There is a broadly defined elliptical nodal belt from which nodal spines radiate. The edge thus performs a fluttering type of vibration at the same time that the central portion is performing a see-saw motion about two axes at right angles. Fig. 17, Plate 3, represents one of the most beautiful types of minor vibration modes met with in discs. It forms one of a group of such modes in which the number of nodal bars diminishes with descending order of frequency. The nodal system consists of two sets of nodes conjugate to each other, referred to two focal regions on a diameter. The outer nodal line of nearly circular form is not actually at the extreme edge of the disc; this would, indeed, be impossible. It appears to be at the edge in the photograph because of a slight chamfer which reduces the effective diameter of the interfering surface. Other modifications of this class of vibration have three annular nodes.

Fig. 18, Plate 3, is a fairly complex vibration, exhibiting torsion as well as flexure. It is almost perfectly symmetrical about a line joining the middle points of the two horizontal sides of the photograph.

Summary.

The fringe patterns obtained by interference of continuous light reflected by the surface of a vibrating quartz plate and by a stationary plane surface have been examined. When the quartz plate is at rest the pattern is uniform. When the quartz vibrates the fringes are blurred in all parts of the surface which are in motion, but their appearance suffers no change where the surface of the quartz plate is motionless. It has been found that at the natural frequency corresponding to the thickness of the plates some portions of the surface do not take part in the vibration and that the amplitude of the vibrating portions varies very much from point to point of the surface.

The interferometer has also been used with intermittent light for stroboscopic examination, which has revealed that the motion of the vibrating plate is non-uniform not only in amplitude, but also in phase, which may differ by as much as 180° between various parts of the surface.

Models in relief of the configuration of various vibrating surfaces at particular instants have been made and photographed.

APPENDIX.

Reproduction of the Vibrating Surfaces in Relief.

An extension of the interferometer method consists in reproducing the vibrating surface in relief from the interference pattern obtained. This development, which is believed to be novel, will be briefly described.

Photographs of vibrating quartz plates taken with intermittent light, such as figs. 2c, Plate 1, 10b, Plate 2, 11b, Plate 2, virtually constitute contour maps of the surfaces at the instant of maximum amplitude of vibration, and the set of curved interference lines represents the traces of parallel equidistant planes on the slightly inclined and distorted quartz surface.

If an enlargement of such an interference contour map be placed under the pointer of a pantographic engraving machine, and at the same time a piece of easily carved material is set up under the engraving tool, it can be arranged, by adjustment of the angle of the upper surface of the material, that the tool carves out the corresponding ledge in the surface when the pointer is moved along a band of the interference diagram. The amount of relief obtained is dependent upon the angle of inclination at which the material is set up in the engraving machine. The photographic enlargement of the vibrating plate is set up on the table of the machine so that the general direction of the

interference bands is parallel with the trace which the surface to be carved makes with the horizontal plane. The pointer is run along the first band with the tool set to carve out the corresponding ledge on the surface. The surface is then moved up the multiple of the half wave-length corresponding to its slope, and the next line is engraved. The whole surface is thus carved out, moving it up on the machine step by step in proceeding from line to line on the photographs.

This principle is of general application where it is required to reproduce in relief, to any scale of magnification, the shape of a surface of which an interference pattern can be obtained, and will no doubt prove useful in other branches of measurement.

The first photograph, fig. 19, Plate 3, corresponds to figs. 2*b* and 2*c*, and needs no explanation. Figs. 20 and 21, Plate 3, correspond to two different modes of vibration of the same disc, represented by figs. 10 and 3*b*, respectively. The first mode is mainly flexural and has a fairly low frequency, whereas in fig. 21 the frequency corresponds to the thickness dimension. This mode is typical of a disc vibrating at a frequency corresponding with its thickness. The resultant motion appears to be made up of overtone compressional vibrations in two directions in the plane of the disc, together with a pure compressional vibration perpendicular to the plane of the disc.

Fig. 22, corresponding to figs. 13*a* and 13*b*, illustrates a very complicated type of vibration, partly flexural and partly torsional. There are three axes along which but little vibration occurs; these are parallel to the diagonal from the upper left-hand corner to the lower right-hand corner, and form "backbones" about which torsional oscillations occur.

Theory of Electric Charge.

By Sir ARTHUR EDDINGTON, F.R.S.

(Received June 10, 1932.)

§ 1. In two earlier papers I have partially developed a theory of the fine structure constant $hc/2\pi e^2$ as arising from the interchangeability of electrons. This theory is, I think, completed in the present paper.

It is agreed that two electrons are indistinguishable and that statistical formulæ must be modified to take account of their interchangeability; but owing to the fact that interchangeability is not easily expressible in terms of *continuous analysis*, its consequences have not hitherto been fully traced. This investigation aims at filling the gap. Thus I am not concerned to invent new hypotheses but to work out the consequences of an old one. Wave mechanics has been successful in replacing "jumps" by continuous analysis, and there seems no reason why it should not prove equally successful with the jump of interchange. Proceeding in this way I find that interchangeability reduces to a term in the wave equation which can be identified with the term ordinarily attributed to the electrostatic and electromagnetic energy of the two charges; and that the coefficient of this term, ordinarily called the fine structure constant, is the integer 137.

It would be absurd to make hypotheses to "explain" indistinguishability, and the many-dimensional geometrical construction used to convert it into a form in which it can be introduced into the differential equations of wave mechanics is not a hypothesis. It is a method of analytical treatment—perhaps not the simplest method possible, but the only one yet devised.

The matrix coefficient of the energy term, given in my earlier papers as $E_1E_1' + E_2E_2' + E_3E_3' + E_4E_4'$ was incorrect. The amended coefficient is found to reduce in special cases to the expression $\frac{1}{2}\{1 + (\sigma, \sigma')\}$ given by Dirac and others. The main point on which my second paper went wrong was in associating interchange with a rotation whose arm was in the E_5 direction; when I realised that the rotation was in the $(P + i)$ direction progress became rapid.

It is necessary to assume a knowledge of the ψ -tensor calculus, and of the practical manipulation of the E-symbols.†

† The earlier papers of the series are in 'Proc. Roy. Soc.,' A, : I, vol. 121, p. 524 (1928); II, vol. 122, p. 358 (1929); III, vol. 126, p. 696 (1930); IV, vol. 133, p. 311 (1931); V, vol. 133, p. 605 (1931); VI, vol. 134, p. 524 (1931). Also VII, 'J. Lond. Math. Soc.,' vol. 7, p. 58 (1931). For the ψ -tensor calculus reference should be made to III and IV. The properties of the E-symbols are given in IV, but are more systematically and fully treated in VII. The earlier papers on the fine structure constant are II and III. Papers V and VI (on the cosmical constant, and the mass of a proton) are referred to only casually.

§ 2. *Conservation of Probability.*

Following Dirac, we regard a "state" of a system as extended over all time. The state is described by a distribution of probability over various possible configurations of the system. The distribution will in general vary with the time, but the integrated probability of all the configurations at any moment of t must be independent of t ; for the idea of a "state" is that a definite probability can be attributed to the state as a whole.

When the configurations are represented by the co-ordinates of a point in generalised space, this conservation of probability can be expressed by introducing a vector \mathbf{j} whose divergence vanishes. The component of \mathbf{j} in the time direction is then the density of probability, and the other components give the density of its flux.

The axiom of wave mechanics is that \mathbf{j} is a product $\psi\psi^*$. Equations determining ψ , ψ^* are called wave equations, and their general form is determined by the condition that they must lead to $\text{div } \mathbf{j} = 0$. Also in order that \mathbf{j} may be a vector, a wave equation must be a tensor equation of the ψ -tensor calculus, *i.e.*, the calculus of quantities whose *products* yield space vectors. This calculus is developed in Papers I, III, IV, and I employ the "complete set" of symbols E_μ as defined in Paper IV. A symbol in clarendon type will denote a complete space-vector. For a single charge the complete infinitesimal space-vector $(dx_1, dx_2, \dots, dx_{16})$ is expressed invariantly by

$$dx = \sum_1^{16} E_\mu dx_\mu. \quad (2.1)$$

The ordinary space vectors of elementary theory (both 4-vectors and 6-vectors) are portions of these complete vectors.

We use two divergence operators denoted by

$$\nabla = \Sigma E_\mu (\partial/\partial x_\mu), \quad \nabla^* = \Sigma (\delta/\delta x_\mu) E_\mu, \quad (2.2)$$

where $\delta/\delta x_\mu$ denotes $\partial/\partial x_\mu$ written *after* its operand. The components of $\psi\psi^*$ are given by (Paper IV, equation (3.4))

$$j_\mu = -\frac{1}{4}\psi^*E_\mu\psi. \quad (2.3)$$

Then

$$\frac{\partial j_\mu}{\partial x_\mu} = -\frac{1}{4}\left\{\psi^*E_\mu\frac{\partial}{\partial x_\mu}\psi + \psi^*\frac{\delta}{\delta x_\mu}E_\mu\psi\right\},$$

so that

$$\text{div } \mathbf{j} = \Sigma \frac{\partial j_\mu}{\partial x_\mu} = -\frac{1}{4}\{\psi^*\nabla\psi + \psi^*\nabla^*\psi\}. \quad (2.4)$$

Let the result of the differentiation be

$$\nabla \psi = \mathbf{m} \psi, \quad \psi^* \nabla^* = \psi^* \mathbf{m}^*. \quad (2.5)$$

Then \mathbf{m} , \mathbf{m}^* , will be space-vectors (in general functions of the co-ordinates); and (2.4) becomes

$$\operatorname{div} \mathbf{j} = -\frac{1}{2} \psi^* (\mathbf{m} + \mathbf{m}^*) \psi. \quad (2.6)$$

We ensure the vanishing of $\operatorname{div} \mathbf{j}$ by taking $\mathbf{m} + \mathbf{m}^* = 0$. Then (2.5) gives the wave equations

$$(\nabla - \mathbf{m}) \psi = 0, \quad \psi^* (\nabla^* + \mathbf{m}) = 0. \quad (2.7)$$

The general condition for the vanishing of $\operatorname{div} \mathbf{j}$ is that the scalar product $\mathbf{j} (\mathbf{m} + \mathbf{m}^*) = 0$. This follows from (2.6) and (2.3). I do not think it would lead to greater generality in physical applications if we used this condition instead of $\mathbf{m} + \mathbf{m}^* = 0$, but in any case we are not here concerned with the theoretical origin of \mathbf{m} , and there is no need to pursue the question.

If we confine attention to solutions of (2.7) which are functions of four rectangular co-ordinates only, viz., $\psi(x_1, x_2, x_3, x_4)$, the equations reduce to

$$\left. \begin{aligned} \left(E_1 \frac{\partial}{\partial x_1} + E_2 \frac{\partial}{\partial x_2} + E_3 \frac{\partial}{\partial x_3} + E_4 \frac{\partial}{\partial x_4} - \mathbf{m} \right) \psi &= 0 \\ \psi^* \left(\frac{\delta}{\delta x_1} E_1 + \frac{\delta}{\delta x_2} E_2 + \frac{\delta}{\delta x_3} E_3 + \frac{\delta}{\delta x_4} E_4 + \mathbf{m} \right) &= 0 \end{aligned} \right\}, \quad (2.8)$$

which are the familiar Dirac equations for an electron.

When the system described by the wave equation is abstracted from the rest of the universe, the interaction of the latter on it is replaced by the interaction of a macroscopic field which is incorporated into the system. It is this field which is described by the vector \mathbf{m} . I have elsewhere (Papers V, VI) begun to develop the theory of \mathbf{m} as arising in this way; but here we only want to know the practical interpretation of the different components of \mathbf{m} ; this is easily found since (2.8) is identical with the equations of current quantum theory. When the field is a pure inertial field, uniform and therefore without gravitation, \mathbf{m} is a scalar constant m called the proper mass. When there is an electromagnetic field, $\mathbf{m} = m + E_1 \kappa_1 + E_2 \kappa_2 + E_3 \kappa_3 + E_4 \kappa_4$, where $(\kappa_1, \kappa_2, \kappa_3, \kappa_4)$ is the electromagnetic potential measured in an appropriate unit. We shall here take the case of a pure inertial field so that $\mathbf{m} = m$.

§ 3. General Transformations.

A linear function of the symbols of a complete set with algebraic coefficients is called a complete number. It is characteristic of complete sets that the

products of complete numbers are complete numbers; the quotients are also complete numbers, except that division by singular complete numbers is excluded in the same way that division by 0 is excluded in algebra. A complete number A is singular if there exists no finite complete number B such that $AB = 1$.

Space-vectors are complete numbers which have in addition the vector transformation property. The most general relativity transformation of a space-vector A is $A' = kAk^{-1}$, where k is any non-singular complete number. Applying the same transformation to a product we have

$$A'B' = kAk^{-1}kBk^{-1} = kABk^{-1}, \quad (3.1)$$

so that the product besides being a complete number obeys the vector transformation law. Thus the products of space-vectors are space-vectors.

The most general transformations of ψ , ψ^* are

$$\psi' = k_1\psi, \quad \psi^{*'} = \psi^*k_2,$$

where k_1 , k_2 are complete numbers; and the corresponding transformation of j is $j' = k_1jk_2$. We are thus led to consider transformations of space-vectors more general than the relativity transformation. Owing to the non-commutative properties of complete numbers the general theory is restricted to infinitesimal transformations. We can write the most general infinitesimal transformation of ψ , ψ^* in the form

$$\psi' = e^{\theta_1 + \theta_2} \psi, \quad \psi^{*'} = \psi^* e^{\theta_1 - \theta_2}, \quad j' = e^{\theta_1 + \theta_2} j e^{\theta_1 - \theta_2},$$

where θ_1 , θ_2 are infinitesimal complete numbers. It therefore falls into two parts, the θ_2 transformation being the relativity transformation kjk^{-1} , while θ_1 gives a new kind of transformation kjk which we shall call a *strain*.

The transformation $k()k^{-1}$ is relativistic. This follows because if it is applied to the E_μ (instead of to the coefficients of the E_μ) it gives new symbols E'_μ which form a complete set intrinsically equivalent to the original set. A strain transformation is not relativistic; if applied to the E_μ it gives new symbols E'_μ which have different commutative properties and do not form a complete set.

We here use relativity in the sense of Einstein's original theory, i.e., restricted to rotations and Lorentz transformations which do not change the metric of the frame of reference. The introduction of strain transformations is a partial substitute for Einstein's more general theory, for they are equivalent to a change of metric (or intrinsic character) of the reference frame E_μ . We

do not actually handle them that way, and the E_μ in our formulæ always denote a complete set with normal metric.†

By describing a quantity as a space-vector we specify its behaviour in regard to vector transformations, or as we now call them relativity transformations, but not in regard to strains. The strain is supposed to be applied to all displacement vectors in the space, and (as in general relativity) we shall call any vector which is transformed by strain in the same way as a displacement vector a *contravariant vector*. The probability vector j will be taken to be a contravariant vector. This is in keeping with its definition as a flux. Strains and relativity transformations are conveniently described by stating the transformations of ψ and ψ^* .

We consider only transformations which are uniform over the whole vector field. They are therefore applicable to co-ordinates as well as to differentials of co-ordinates.

§ 4. Theory of Rotations.

An infinitesimal rotation can be specified by a displacement vector $d\mathbf{x}$ and an arm \mathbf{r} ; the latter is a vector drawn from the centre of rotation to the point displaced. We define the rotation as the transformation

$$\psi' = e^{\frac{1}{2} d\mathbf{x} \cdot \mathbf{r}^{-1}} \psi, \quad \psi^{*'} = \psi^* e^{\frac{1}{2} \mathbf{r}^{-1} \cdot d\mathbf{x}}, \quad (4.1)$$

applied to the contravariant vectors descriptive of that which undergoes the rotation. The angle of rotation is $d\theta = d\mathbf{x} \cdot \mathbf{r}^{-1}$.

For example, consider three-dimensional space with the origin at the centre of rotation, so that

$$\mathbf{r} = E_1 x_1 + E_2 x_2 + E_3 x_3, \quad d\mathbf{x} = E_1 dx_1 + E_2 dx_2 + E_3 dx_3 \quad (4.2)$$

Then

$$\begin{aligned} \mathbf{r} d\mathbf{x} &= -x_1 dx_1 - x_2 dx_2 - x_3 dx_3 + E_{23}(x_2 dx_3 - x_3 dx_2) + \dots \\ &= -r dr + r^2 (E_{23} d\theta_{23} + E_{31} d\theta_{31} + E_{12} d\theta_{12}), \end{aligned} \quad (4.3)$$

where $d\theta_{23}$, $d\theta_{31}$, $d\theta_{12}$ are the ordinary components of angular rotation. By (4.2), $r^2 = -\mathbf{r}^2$; hence $\mathbf{r}^{-1} = -\mathbf{r}/r^2$. Hence by (4.3)

$$\mathbf{r}^{-1} \cdot d\mathbf{x} = d(\log r) - E_{23} d\theta_{23} - E_{31} d\theta_{31} - E_{12} d\theta_{12}. \quad (4.41)$$

Similarly

$$d\mathbf{x} \cdot \mathbf{r}^{-1} = d(\log r) + E_{23} d\theta_{23} + E_{31} d\theta_{31} + E_{12} d\theta_{12}. \quad (4.42)$$

† The purpose of Einstein's more general transformations is to provide a means of dealing with space of irregular curvature in which no uniform metric can exist. We deal only with flat space in which a normal metric is always available.

Inserting these values in (4.1) the part of the rotation involving $d\theta_{12}$ is

$$\psi' = e^{\frac{1}{2} E_{12} d\theta_{12}} \psi, \quad \psi^{*'} = \psi^* e^{-\frac{1}{2} E_{12} d\theta_{12}}. \quad (4.5)$$

Hence by (2.3)

$$\begin{aligned} j_1' &= -\frac{1}{2} \psi^{*'} E_1 \psi' = -\frac{1}{2} \psi^* (e^{-\frac{1}{2} E_{12} d\theta_{12}} E_1 e^{\frac{1}{2} E_{12} d\theta_{12}}) \psi \\ &= -\frac{1}{2} \psi^* E_1 (\cos d\theta_{12} + E_{12} \sin d\theta_{12}) \psi, \text{ since } E_1 \text{ and } E_{12} \text{ anticommute} \\ &= -\frac{1}{2} \psi^* (E_1 \cos d\theta_{12} - E_2 \sin d\theta_{12}) \psi \\ &= j_1 \cos d\theta_{12} - j_2 \sin d\theta_{12}, \end{aligned}$$

which shows that space vectors are rotated through an angle $d\theta_{12}$.

Writing—

$$\log r = i\phi, \quad (4.6)$$

the part of the rotation involving dr is (since $E_{16} = i$)

$$\psi' = e^{i E_{16} d\phi} \psi, \quad \psi^{*'} = \psi^* e^{i E_{16} d\phi}, \quad (4.7)$$

which is a strain transformation.

It will be seen that if in (4.1) we divide dx into two parts, the first perpendicular to (*i.e.*, anticommuting with) r and the second antiperpendicular to (commuting with) r , the first part gives relativity rotations and the second part strains. For this reason we sometimes use the term *antiperpendicular rotation* instead of strain.

Antiperpendicular rotations rotate pairs of antiperpendicular terms in the same way that perpendicular (or relativity) rotations rotate pairs of perpendicular terms. For example, if $k = e^{i E_{12} \theta_{12}}$, the antiperpendicular rotation $\psi' = k\psi$, $\psi^{*'} = \psi^* k$, gives

$$\begin{aligned} -4j_5' &= \psi^{*'} k E_5 k \psi = \psi^* E_5 (\cos \theta_{12} + E_{12} \sin \theta_{12}) \psi \\ &= \psi^* (E_5 \cos \theta_{12} - i E_{34} \sin \theta_{12}) \psi \\ &= -4(j_5 \cos \theta_{12} - ij_{34} \sin \theta_{12}). \end{aligned}$$

Since for real physical vectors j_5 is real and j_{34} is imaginary, this is a real circular rotation of the antiperpendicular components j_5 and ij_{34} . The appearance of i is automatic, and it is a general rule that if in $k = e^{i E_{\mu\nu} \theta_{\mu\nu}}$ we choose $\theta_{\mu\nu}$ (real or imaginary) so that k is real, then k gives real (circular or hyperbolic) perpendicular rotations and also real antiperpendicular rotations.

In current quantum theory it is the practice (which we do not follow) to take ψ and ψ^* to be conjugate complex quantities, so that any transformation factors attached to them are also taken to be conjugate complex quantities.

The perpendicular and antiperpendicular rotations then correspond to the imaginary and real parts of the angular transformations. It would be inconvenient in relativity theory not to know whether a transformation is relativistic or not until we are told whether the symbol denotes a real or imaginary quantity; and the convention is difficult to use in Dirac's theory which leads to 16 different symbols for the square root of -1 some of which are purely real. Moreover, the convention, if consistently carried out, makes all the strains real and the relativity rotations imaginary. We therefore do not follow the practice; but it has seemed of interest to note the way in which antiperpendicular rotations or strains have been introduced by earlier writers.

When \mathbf{r} is a general vector the relation of \mathbf{r}^{-1} and \mathbf{r} is not so simple as in the above example in which it contained only mutually perpendicular components. We can, however, always find \mathbf{r}^{-1} from \mathbf{r} by solving the vector equation $\mathbf{r}^{-1}\mathbf{r} = 1$. Unless \mathbf{r} is a singular vector (a generalisation of $r = 0$ in algebra), \mathbf{r}^{-1} exists and is unique and commutes with \mathbf{r} , and anything which commutes with \mathbf{r} commutes with \mathbf{r}^{-1} .

§ 5. Covariant Derivatives.

A distribution of probability over the four dimensional domain x_1, x_2, x_3, x_4 , the other components of x being fixed, will be called a *sub-state*. It does not necessarily form a conserved distribution since there may be flux in directions not included in the sub-state. Consider a rotational or strain transformation

$$\psi(x, \theta_\mu) = e^{iE_\mu\theta_\mu} \psi(x, 0),$$

where x denotes x_1, x_2, x_3, x_4 , collectively. Then θ_μ can be regarded as a parameter of the sub-state. By taking $\mu = 1, 2, \dots, 16$, and using both rotations and strains, we classify 32 different ways† in which a sub-state (*i.e.*, a probability distribution) can be varied infinitesimally. The classification does not apply directly to finite changes, because in general these do not commute and consequently cannot be compounded.

The 32 modes of variation would be redundant if we had only to deal with single ψ -vectors which admit only 8 independent variations; but the method of wave mechanics is to obtain general solutions by superposing elementary solutions of the wave equation, so that we are concerned with transformations not as they affect a particular ψ but any number of ψ 's simultaneously.

† Strictly 30 different ways, because the E_{16} transformations which are algebraic are considered to change not the sub-state but the probability attached to the sub-state as a whole.

A much greater variety of adjacent states would be obtained by admitting non-uniform transformations as in Einstein's general theory. Our object, however, is not to pursue the most general development conceivable, but to obtain an extension of ordinary four-dimensional conservation sufficient to handle the phenomenon of interchangeability of charges which is the purpose of this paper. A simple extension is to take a single θ_μ , and consider a state made up of all the sub-states corresponding to different values of θ_μ . Thus θ_μ which was a parameter of the sub-state becomes a co-ordinate of the state, and the derivative of the flux in the θ_μ direction must be included in the divergence of j .

We need not attach equal probability to each of the sub-states; thus the probability distribution of the state can be taken as

$$\psi(x, \theta_\mu) = e^\alpha \cdot e^{i\theta_\mu \theta_\mu} \psi(x, \theta), \quad (5.1)$$

where α is an algebraic factor. It must be algebraic, because a general symbolic factor would alter the sub-state.

By Taylor's theorem

$$\psi(x, \theta_\mu) \equiv e^{\theta_\mu \partial / \partial \theta_\mu} \psi(x, \theta).$$

Hence

$$e^{\theta_\mu (\partial / \partial \theta_\mu - \frac{1}{2} E_\mu)} \psi = e^\alpha \psi. \quad (5.2)$$

We write

$$M_\mu = \partial / \partial \theta_\mu - \frac{1}{2} E_\mu, \quad (5.3)$$

and call M_μ the *covariant derivative*. When θ_μ is the angle of a strain M_μ applies also to ψ^* ; when θ_μ is a relativity rotation, it is replaced by $M^*_\mu = \partial / \partial \theta_\mu + \frac{1}{2} E_\mu$.

The term covariant derivative is used in the same sense as in the tensor calculus; for progress along the co-ordinate θ_μ in the state involves a rotation or strain of the co-ordinate directions x as compared with their original directions, and part of the change of ψ is the effect of this co-ordinate change on a contravariant vector. The covariant derivative separates out the "real change" between the point θ_μ and $\theta_\mu + d\theta_\mu$ in the state, discarding what is merely due to the altered scheme of reference applicable at the latter point.

It is clear that, as in general relativity, the covariant derivative must be used in the divergence. The simple derivative, as used in (2.4), applies when we are using a rectangular co-ordinate in the direction considered; so that progress along it involves no change in the mode of resolution of ψ . If then we employ an angular co-ordinate we must use the covariant derivative in

order not to bring in the spurious change due to the altered directions of resolution.

In order that $\psi(x, \theta_\mu)$ may be single-valued,† the coefficient e^a must be a periodic function of θ_μ with period 2π . We therefore analyse it into its Fourier components $e^{in\theta_\mu}$, and analyse the general state into elementary states corresponding to the different values of n . The integer n is then a parameter of the elementary state. Thus, for an elementary state, we have by (5.2)

$$e^{\theta_\mu M_\mu} \psi = e^{in\theta_\mu} \psi, \quad (5.4)$$

and we have therefore

$$M_\mu = in, \quad (n = 0, \pm 1, \pm 2, \dots). \quad (5.5)$$

The possibility of M having other than algebraic values is excluded by the afore-mentioned condition that the transformation has to apply to states composed of the sum of a number of ψ 's.

In quantum theory differentiation operators are usually called momenta, and M_μ is accordingly called angular momentum. By (5.5) angular momentum is quantised. The "spin term" $-\frac{1}{2}E_\mu$ in the angular momentum is due to the operator being a covariant derivative.

§ 6. Transformation to Polar Co-ordinates.

In § 4 we introduced angular elements $dx \cdot r^{-1}$, $r^{-1} dx$ and we now consider the analogous angular divergences $r\nabla$, ∇^*r . The multiplication is similar. Let

$$r = E_1x_1 + E_2x_2 + E_3x_3 = E_r r, \quad (6.1)$$

so that $E_r^2 = -1$. Multiply the equations of (2.8) by initial and final E_r respectively; we obtain

$$\left(E_{16} \frac{\partial}{r\partial\phi} + E_{23} \frac{\partial}{r\partial\theta_{23}} + E_{31} \frac{\partial}{r\partial\theta_{31}} + E_{12} \frac{\partial}{r\partial\theta_{12}} + E_{r4} \frac{\partial}{\partial x_4} - E_r m \right) \psi = 0 \quad (6.21)$$

$$\psi^* \left(\frac{\delta}{r\delta\phi} E_{16} - \frac{\delta}{r\delta\theta_{23}} E_{23} - \frac{\delta}{r\delta\theta_{31}} E_{31} - \frac{\delta}{r\delta\theta_{12}} E_{12} - \frac{\delta}{\delta x_4} E_{r4} + E_r m \right) = 0. \quad (6.22)$$

† Strictly, in order that ψ may change to $-\psi$ after a revolution. The problem of one charge is highly abstract, and the rotations here considered do not begin to have a physical interpretation until we reach the problem of two charges. Rotations involve a singularity at the centre of rotation, which is allowed provisionally on the understanding that it will be replaced by a second charge; the combined ψ of the two charges will then be single-valued, each contributing a factor -1 after a revolution 2π . Unless one is prepared to plunge directly into a highly complicated problem, it is difficult to avoid provisional adaptations of this kind.

Since the rotations θ_{23} , θ_{31} , θ_{12} anticommute only one of them, or one combination of them, can be used as a finite co-ordinate. The same limitation occurs in elementary geometry where only one such rotation can be used as a polar co-ordinate ϕ . Since E_{23} , E_{31} , E_{12} are mutually perpendicular, the three terms corresponding to them combine into a single term $E_\theta \partial/\partial\theta$, and θ is evidently in a plane containing r . Thus (6.21) reduces to

$$\left(E_{16} \frac{\partial}{r \partial \phi} + E_\theta \frac{\partial}{r \partial \theta} + E_{r4} \frac{\partial}{\partial x_4} - E_r m \right) \psi = 0. \quad (6.3)$$

Since we have not modified ψ it is still referred to constant rectangular directions. In order to use θ as an angular co-ordinate of a state, like θ_μ in § 5, ψ must be transformed. Equation (6.3) will, however, continue to hold if we substitute the covariant derivative M_θ for the ordinary derivative $\partial/\partial\theta$. By § 5, M_θ is quantised and equal to in . Hence (6.3) becomes

$$\left(-\frac{\partial}{\partial r} + E_\theta \frac{in}{r} + E_{r4} \frac{\partial}{\partial x_4} - E_r m \right) \psi = 0. \quad (6.4)$$

This is equivalent to Dirac's wave equation ("Quantum Mechanics," p. 252, (I. 4)); he gives different matrix coefficients but they satisfy the same commutative relations as ours. (E_θ is perpendicular to E_r and antiperpendicular to E_{r4} .)

If we use the angular co-ordinate ϕ instead of r , we must also replace $\partial/\partial\phi$ in (6.3) by the covariant derivative ($\partial/\partial\phi - \frac{1}{2}E_{16}$). This gives an extra term $1/2r$ in (6.4). If ψ is a solution of (6.4) the solution of the equation including the extra term is $r^\frac{1}{2}\psi$. This may also be written $e^{\frac{1}{2}E_{16}\phi}\psi$, and is the transformation already indicated in (4.7). The question of quantisation of M_θ does not arise since the real rotation is hyperbolic. Consequently, we usually retain the linear co-ordinate r .

We have deferred mention of an important point. If (6.21) and (6.22) are multiplied respectively by ψ^* and ψ and added, we obtain $\partial j_{16}/\partial\phi = 0$; so that there is no flux along r . This means that our equations relate to a trivial problem. The only directions of flow contemplated in (6.2) are r , θ_{12} , θ_{23} , θ_{31} , and since the last three are *relativity* rotations which cannot change the probability, the change of flow is actually restricted to the direction r . The equation gives the obvious result that under such restrictions the flow is constant. Equation (6.4) is illusory, because the quantum condition that n must be an integer or zero is further restricted by a relativity condition that n is zero.

The triviality is removed by substituting for the rotational angle θ a corresponding strain angle, and for the relative time x_4 corresponding to a Lorentz transformation a proper time s corresponding to a strain transformation. The substitution of a non-relative time (which in current quantum theory is effected by substituting a real for an imaginary variable) does no harm at this stage, for we have already eliminated the possibility of Lorentz transformations by taking the three-dimensional vector r to be definite. If it were desired to retain relative time a four-dimensional vector should have been used.

The rotations are changed to strains by altering the sign of the corresponding terms in ψ^* . The equation corresponding to (6.3) is then

$$\psi^* \left(\frac{\delta}{r\delta\phi} E_{1\theta} + \frac{\delta}{r\delta\theta} E_\phi + \frac{\delta}{\delta s} E_{r4} + E_{rm} \right) = 0 \quad (6.5)$$

and s replaces x_4 in (6.3). Multiplying by ψ^* and ψ and adding, we obtain an equation which is no longer trivial, and allows flux of probability in the directions r , θ and s . There is no longer an overriding condition $n = 0$; so that (6.4) is rehabilitated.

It will be seen that the polar equations (6.3), (6.4), (6.5) are not derived directly from the Cartesian equations (2.8), but by considering conserved flow of probability in a three-dimensional continuum r , θ , s , which has only one dimension in common† with the continuum x_1, x_2, x_3, x_4 . The change of time might have been avoided if we had adopted at the outset an independent variable x_{45} antiperpendicular to three-dimensional space, so that $E_{45}\partial/\partial x_{45}$ replaced $E_4\partial/\partial x_4$ in (2.8). Since E_r commutes with E_{45} , the term $E_{r45}\partial/\partial x_{45}$ appears with the same sign in (6.21) and (6.22).

§ 7. The Symbol P .

For two charges we introduce two complete sets of symbols E_μ and F_μ . Every E_μ commutes with every F_μ . It was pointed out by G. Temple that there exists an operator

$$P = \frac{1}{4i} \sum_1^{16} E_\mu F_\mu \quad (7.1)$$

which satisfies $P^2 = -1$, and transforms one set into the other according to the law

$$F_\mu = -PE_\mu P. \quad (7.2)$$

† In Dirac's language only one dimension can be represented by a diagonal matrix at the same time as the Cartesian co-ordinates.

It follows that

$$PF_\mu = E_\mu P, \quad PE_\mu = F_\mu P, \quad PE_\mu F_\nu = E_\nu F_\mu P. \quad (7.3)$$

The most general displacement is $d\mathbf{x} = \sum_{\mu} \sum_{\nu} E_\mu F_\nu dx_{\mu\nu}$, so that the generalised space is 256-dimensional and complete numbers of the new series each contain 256 terms. We divide $d\mathbf{x}$ into two parts

$$d\mathbf{x} = d\mathbf{x}^s + d\mathbf{x}^a = \sum_1^{136} \gamma_\mu dx_\mu^s + \sum_1^{120} \zeta_\mu dx_\mu^a, \quad (7.4)$$

where

$$\gamma_{\mu\nu} = \frac{1}{2} (E_\mu F_\nu + E_\nu F_\mu), \quad \zeta_{\mu\nu} = \frac{1}{2} (E_\mu F_\nu - E_\nu F_\mu). \quad (7.5)$$

A single suffix notation for the γ 's and ζ 's is often sufficient and is used in (7.4).

By (7.3) and (7.5)

$$P\gamma_{\mu\nu} = \gamma_{\mu\nu} P, \quad P\zeta_{\mu\nu} = -\zeta_{\mu\nu} P. \quad (7.6)$$

Thus P commutes with $d\mathbf{x}^s$ and anticommutes with $d\mathbf{x}^a$. The vector $\mathbf{j} = \sum (\gamma_\mu j_\mu^s + \zeta_\mu j_\mu^a)$ is likewise divisible into two parts. By a simple generalisation of (2.3) its components are

$$j_\mu^s = \frac{1}{16} \psi^* \gamma_\mu \psi, \quad j_\mu^a = \frac{1}{16} \psi^* \zeta_\mu \psi. \quad (7.7)$$

§ 8. *Interchange of the Charges.*

We denote the co-ordinates of the two charges in space-time collectively by x, x' so that the combination gives a point (x, x') in an 8-dimensional continuum. It results from the interchangeability of the two charges that what is physically the same point or configuration A is represented analytically twice over at (x, x') and at (x', x) . Actually there can be only one probability of the configuration A ; we have to divide this probability and assign one part to (x, x') and the other to (x', x) . This division is necessarily arbitrary. If our analytical equations show that a quantity of probability has jumped from (x, x') to (x', x) the jump means nothing physically and there is no real breach of continuity.

Thus the problem of conserved flow of probability for two charges is not a simple extension to eight dimensions of the equation of continuity $\text{div } \mathbf{j} = 0$. In four dimensions the rule was that probability must flow continuously; while for eight dimensions the rule is that probability must flow continuously or it may jump from one point to a complementary point. Naturally, it is an embarrassment to continuous analysis to allow a jump of any kind, even though,

as here, each point has only one complementary point to which a jump may be allowed.

The only successful method of dealing with this is to build a *reservoir* of probability outside the eight dimensions of the state. Probability can flow continuously from the state into the reservoir and vice versa. We then govern the continuous flow in the reservoir by the condition (which will be found to be a quantising condition) that the amount flowing from the reservoir into the state at the point (x, x') at any moment is always equal to the amount flowing from the state into the reservoir at the point (x', x) .

The reservoir is a fictitious analytical construction for the purpose of replacing jumps by continuous flow. If it were not for the jumps there would be no object in introducing dimensions outside the actual dimensions of space-time. The jumps in their turn are a fiction necessitated by our replacing actual microscopic space which must be an abstraction of the relations of indistinguishable particles by a geometrical space correlated to the physics of distinguishable macroscopic bodies.

§ 9. Continuous Interchange.

When E_μ and F_μ are represented by matrices the wave function is denoted by $\psi_{\alpha\beta}(x, x')$. Then E_μ and F_μ denote the *same* matrix, but E_μ forms chain products with the first suffix and F_μ with the second. We regard x, α as co-ordinates of charge No. 1, and x', β as co-ordinates of No. 2, α and β being "spin co-ordinates."

It can be shown that

$$P\psi_{\alpha\beta} = i\psi_{\beta\alpha}, \quad \psi^*_{\alpha\beta}P = i\psi^*_{\beta\alpha}. \quad (9.1)$$

For by (7.2)

$$F_\mu\psi_{\alpha\beta} = (iP)E_\mu(iP)\psi_{\alpha\beta}, \quad (9.2)$$

which signifies that instead of performing a certain operation F_μ on the *second* suffix, we can make the change in ψ indicated by iP , perform the same operation on the *first* suffix, and then change ψ back again. (The change iP is self-reciprocal since $(iP)^2 = 1$.) Clearly the change indicated by iP must be reversal of the two suffixes. It might in addition include some other self-reciprocal change Q which would be innocuous provided that Q commutes with E_μ . There is an equation similar to (9.2) for every E_μ and F_μ , so that Q would have to commute with all of them; Q must therefore be algebraic.†

† An *algebraic number* is defined to be a symbol which commutes with every symbol in the calculus.

The only self-reciprocal algebraic operators are ± 1 . By a more direct proof I find that the sign is as stated in (9.1). That the sign attached to i must be the same for ψ and ψ^* is verified by inner multiplication

$$\psi^*_{\alpha\beta} P^2 \psi_{\alpha\beta} = \psi^*_{\beta\alpha} i^2 \psi_{\beta\alpha}.$$

In the inner products α and β are dummies, so that $\psi^*_{\alpha\beta} \psi_{\alpha\beta} \equiv \psi^*_{\beta\alpha} \psi_{\beta\alpha}$.

Since $e^{i\pi P} = \cos \frac{1}{2}\pi + P \sin \frac{1}{2}\pi = P$, (9.1) can also be written

$$e^{i\pi P} \psi_{\alpha\beta} = i \psi_{\beta\alpha}, \quad \psi^*_{\alpha\beta} e^{i\pi P} = i \psi^*_{\beta\alpha}. \quad (9.3)$$

Consider the rotation

$$k = e^{i(P+i)\chi}, \quad k^{-1} = e^{-i(P+i)\chi}. \quad (9.4)$$

Take first the relativity rotation, so that $\psi \rightarrow k\psi$, $\psi^* \rightarrow \psi^* k^{-1}$, $j \rightarrow kjk^{-1}$. When $\chi = \pi$ this gives

$$\psi_{\alpha\beta} \rightarrow -\psi_{\beta\alpha} \quad \psi^*_{\alpha\beta} \rightarrow -\psi^*_{\beta\alpha}.$$

Further the transformation applied to the vector giving the position of the two charges relative to the origin, gives

$$\begin{aligned} \Sigma (E_\mu x_\mu + F_\mu x'_\mu) &\rightarrow \Sigma e^{i\pi(P+i)} (E_\mu x_\mu + F_\mu x'_\mu) e^{-i\pi(P+i)} \\ &= -\Sigma P (E_\mu x_\mu + F_\mu x'_\mu) P \\ &= \Sigma (E_\mu x'_\mu + F_\mu x_\mu) \quad \text{by (7.2).} \end{aligned}$$

The complete relativity transformation for $\chi = \pi$ is therefore

$$\psi_{\alpha\beta}(x, x') \rightarrow -\psi_{\beta\alpha}(x', x), \quad \psi^*_{\alpha\beta}(x, x') \rightarrow -\psi^*_{\beta\alpha}(x', x), \quad (9.5)$$

that is to say it interchanges the co-ordinates (both positional and spin) of the two charges and in addition reverses the signs of ψ , ψ^* .

The strain transformation, $\psi \rightarrow k\psi$, $\psi^* \rightarrow \psi^* k$, differs from the relativity rotation at intermediate values of χ but meets it again at $\chi = \pi$, and for that value gives just the same result. This is evident because when $\chi = \pi$, $k = k^{-1}$ by (9.4).

The relativity transformation is of no great interest since there can be no real change of probability in that direction as was seen in § 6; but the strain transformation is of great importance because it gives a route (through the "reservoir") along which probability can flow continuously from (x, x') to (x', x) . We use the strain transformation in the way explained in § 5. Let

$$\psi_{\alpha\beta}(x, x', \chi) = e^{\frac{i}{2}(P+i)\chi} \psi_{\alpha\beta}(x, x', 0), \quad \psi^*_{\alpha\beta}(x, x', \chi) = \psi^*_{\alpha\beta}(x, x', 0) e^{\frac{i}{2}(P+i)\chi}. \quad (9.6)$$

For fixed values of χ , the distribution of ψ , ψ^* over the domain x, x' forms sub-states, χ being the parameter defining a particular sub-state. By turning χ into a co-ordinate we combine the sub-states into a state, the distribution over the nine-fold domain being the state. To adapt this to the physical problem as described in § 8 we have to impose two conditions. Firstly, probability must be conserved in the nine-dimensional state ; this is expressed by the usual wave equations with the inclusion of the term arising from the added dimension. Secondly, if j_χ is the component of \mathbf{j} in the added dimension, the contributions of j_χ to $\text{div } \mathbf{j}$ at (x, x') and (x', x) must be equal and opposite.

§ 10. *Quantisation of χ -Momentum.*

Regarding (9.6) from the point of view of a transformation, as χ increases the co-ordinate frame is transformed by strain so that the point (x, x') changes to a different point in the old frame and ψ undergoes the consequential change of a contravariant vector. When $\chi = \pi$, the result of the strain is given by (9.5), viz., $\psi_{\alpha\beta}(x, x')$ in the strained frame has a value equal to $-\psi_{\beta\alpha}(x', x)$ in the original frame.

The relation between the strained and original co-ordinate frames is that every point (x, x') is changed into its complementary point (x', x) . This holds not only for four dimensions but for every dimension of a complete vector \mathbf{x} . But a co-ordinate frame of space-time does not take cognizance of an order of numbering of points that are referred to it, so that this alteration of order is not a change of space-time frame, though it is a change of the eight-dimensional frame. The physical frames being indistinguishable, we have the result

$$\psi_{\alpha\beta}(x, x') = -\psi_{\beta\alpha}(x', x), \quad (10.1)$$

both being referred to the *same* four-dimensional frame of reference. This is the Fermi-Dirac law of antisymmetry of wave-functions of interchangeable charges.

What we have here determined is the condition that two interchangeable points in four dimensions can be represented for analytical purposes by one point in eight dimensions. Interchange of the points, which is a simple transposition of the arguments of a function without change of co-ordinate system in four dimensions, is a strain of the eight-dimensional co-ordinate frame. The effect of the strain on the factor ψ of a contravariant vector is known ; the condition that transposition of the arguments of ψ without change

of four-dimensional co-ordinate system must give the same result is here found to be that ψ is an antisymmetrical function.

The covariant derivative separates out the "real change" of ψ from the change attributable to change of co-ordinate frame. For the χ co-ordinate we reach different results according as our standpoint is eight-dimensional or four-dimensional. In eight dimensions the factor -1 in (10.1) is part of the effect on ψ of straining the co-ordinate frame, and is not a "real change." In four dimensions there is no change of co-ordinate frame; and after rewriting $\psi_{\alpha\beta}(x, x')$ as $\psi_{\beta\alpha}(x', x)$ in accordance with the renumbering of the points whatever factor is left over is a real change of ψ . Thus the factor -1 is a real change.

We now make the transition from eight dimensions to four dimensions. As we have seen the price of this step is that we introduce the real change of ψ by a factor -1 when χ increases by π ; or equivalently that we introduce the corresponding covariant derivative or angular momentum M_χ .

This step may also be described as the recognition that x and x' refer to the same world and not to disconnected worlds arbitrarily combined to form an eight-dimensional domain. Until the points described by x and x' are in the same world there is no meaning in speaking of a "distance" between them. Thus this step introduces for the first time the physical conception of a distance between two entities. We may expect that the conditions of introducing the conception of distance will automatically introduce the energy and momentum associated with distance, *i.e.*, the electrostatic energy of the entities separated by the distance. This expectation is fulfilled, for we shall find that the electrostatic energy arises from the angular momentum M_χ due to the factor -1 which is introduced by the same step.

Writing the real change in the form e^{xM_χ} as in (5.2), (5.3), we have

$$e^{\pi M_\chi} \psi = -\psi = e^{(2n+1)\pi} \psi. \quad (10.2)$$

Thus $e^{\pi M_\chi}$ can be analysed into terms of period $2\pi(2n+1)$. As in §5 the general state is obtained by superposition of elementary states corresponding to the separate components, and for an elementary state

$$M_\chi = i(2n+1). \quad (10.3)$$

To verify that the condition at the end of §9 is fulfilled, we must calculate j_χ explicitly. Let

$$x = \sum_{\mu=1}^4 (E_\mu x_\mu + F_\mu x'_\mu) = \sum_1^4 \frac{1}{2} (E_\mu + F_\mu) (x_\mu + x'_\mu) + \sum_1^4 \frac{1}{2} (E_\mu - F_\mu) (x_\mu - x'_\mu). \quad (10.4)$$

The point that is being considered is the end of this vector ; the new direction of flow at that point, giving the component j_x , is in the plane of rotation ($P + i$) and at right angles to \mathbf{x} . It is therefore the direction of $d\mathbf{x}$ when a perpendicular rotation ($P + i$) is given.† Since i commutes with all symbols and P commutes with $\frac{1}{2}(\mathbf{E}_\mu + \mathbf{F}_\mu)$, we have

$$\begin{aligned} d\mathbf{x} &= P d\theta \cdot \Sigma \frac{1}{2} (\mathbf{E}_\mu - \mathbf{F}_\mu) (x_\mu - x'_\mu) \\ &= r_{12} d\theta \cdot \frac{1}{2} P (\mathbf{E}_r - \mathbf{F}_r), \end{aligned} \quad (10.5)$$

where r_{12} is the distance between the charges, and

$$\mathbf{E}_r = \Sigma l_\mu \mathbf{E}_\mu, \quad \mathbf{F}_r = \Sigma l_\mu \mathbf{F}_\mu, \quad l_\mu = (x_\mu - x'_\mu)/r_{12}, \quad (10.6)$$

so that $\mathbf{E}_r^2, \mathbf{F}_r^2 = -1$. Hence by (7.7)

$$j_x = \frac{1}{16} \psi^* \frac{1}{2} P (\mathbf{E}_r - \mathbf{F}_r) \psi, \quad (10.7)$$

reckoned on the same scale as the components j_μ^s, j_μ^a in the plane x, x' .

Accordingly in the divergence equation we introduce a new rectangular direction $\frac{1}{2}P(\mathbf{E}_r - \mathbf{F}_r)$, \mathbf{E}_r and \mathbf{F}_r being fixed symbols independent of the point considered. This is the new dimension of flow at the point considered and at any other point in the same line r_{12} , though not at points off this line. The component j_x in this direction is equal and opposite at (x, x') and (x', x) , for

$$\psi^*_{\alpha\beta}(x', x) \mathbf{E}_\mu \psi_{\alpha\beta}(x', x) \equiv \psi^*_{\beta\alpha}(x', x) \mathbf{F}_\mu \psi_{\beta\alpha}(x', x) = \psi^*_{\alpha\beta}(x, x') \mathbf{F}_\mu \psi_{\alpha\beta}(x, x'),$$

so that interchanging the co-ordinates is compensated by interchanging \mathbf{E}_μ and \mathbf{F}_μ in j_x . Hence by (10.7) j_x is reversed in sign. For a uniform strain the covariant derivative of ψ is $M_x \psi$ or $i(2n+1)\psi$ irrespective of the argument of ψ . Thus the covariant derivative $(j_x)_x$ is equal and opposite at the two points ; and what it contributes to the divergence at the one point it takes away at the other.

§ 11. Vector Density.

The product of a number of infinitesimal displacement vectors gives a volume element

$$dv = d\mathbf{x}_{(1)} d\mathbf{x}_{(2)} d\mathbf{x}_{(3)} \dots \quad (11.1)$$

† The direction is given by the perpendicular rotation not by the strain. For example, if we are given a radius in the direction \mathbf{E}_1 and a strain in the \mathbf{E}_{12} plane and wish to find a geometrical construction for the second direction \mathbf{E}_2 defining the strain, it is no use examining how the end of the radius vector moves under the strain or antiperpendicular rotation \mathbf{E}_{12} , because the antiperpendicular rotation leaves perpendicular directions unaltered.

By (3.1) the volume element is a space-vector ; but since it does not transform like a displacement vector under strain, it is not a contravariant vector.

A finite volume or range of displacement can always be resolved into *diagonal* volume elements, so that without loss of generality we can confine attention to volume elements which are diagonal if referred to unstrained co-ordinates. For a single charge the full diagonal volume element is $dv = E_1 dx_1 \cdot E_2 dx_2 \dots E_{16} dx_{16}$. For two charges we write

$$dv = dv^s dv^a \quad (11.21)$$

where

$$dv^s = \gamma_1 dx_1^s \cdot \gamma_2 dx_2^s \dots \gamma_{136} dx_{136}^s, \quad dv^a = \zeta_1 dx_1^a \cdot \zeta_2 dx_2^a \dots \zeta_{120} dx_{120}^a \quad (11.22)$$

For a strain transformation k , we have

$$dv'^s = k\gamma_1 dx_1^s k^2 \gamma_2 dx_2^s k^2 \dots \gamma_{136} dx_{136}^s k.$$

In general this is a highly complicated result, since k will not commute with all the γ 's. But when $k = e^{i\mathbf{F}\cdot\mathbf{x}}$, the result simplifies. Since k commutes with every γ and anticommutes with every ζ by (7.6), we have

$$dv'^s = e^{136\mathbf{F}\cdot\mathbf{x}} dv^s \quad dv'^a = dv^a \quad dv' = e^{136\mathbf{F}\cdot\mathbf{x}} dv. \quad (11.3)$$

The vector \mathbf{j} measures the density of probability, and to obtain a finite probability we must multiply by a volume element and integrate. To preserve the symmetrical relation of ψ and ψ^* to the co-ordinate frame the product must be written in the form

$$dv^{\frac{1}{2}} \mathbf{j} dv^{\frac{1}{2}} = dv^{\frac{1}{2}} \psi \psi^* dv^{\frac{1}{2}}. \quad (11.4)$$

As in the usual tensor calculus we introduce vector densities $\mathfrak{J} = \Psi \Psi^*$ such that

$$dv^{\frac{1}{2}} \psi = dv^{\frac{1}{2}} \Psi \quad \psi^* dv = \Psi^* dv^{\frac{1}{2}}, \quad (11.5)$$

where $dv^{\frac{1}{2}}$ is *invariant*. For a general strain the transformation of $dv^{\frac{1}{2}}$ would require a special investigation. (I am not sure even that simple factorisation of dv into $dv^{\frac{1}{2}} \cdot dv^{\frac{1}{2}}$ would be possible in all cases.) But for the strain $e^{i\mathbf{F}\cdot\mathbf{x}}$ the difficulty of non-commutability does not arise, and we have by (11.3)

$$dv'^{\frac{1}{2}} = e^{68\mathbf{F}\cdot\mathbf{x}} dv^{\frac{1}{2}}.$$

Similarly for the strain $e^{i\mathbf{x}\cdot\mathbf{x}}$, we have $dv'^{\frac{1}{2}} = e^{123i\mathbf{x}\cdot\mathbf{x}} dv^{\frac{1}{2}}$.

Hence the transformation of vector density corresponding to (9.6) is

$$\Psi_{\alpha\beta}(x, x', \chi) = e^{i\mathbf{x}\cdot(137\mathbf{F} + 257\mathbf{i})} \Psi_{\alpha\beta}(x, x', 0). \quad (11.6)$$

The exponential factor becomes algebraic when χ is a multiple of $2\pi/137$.

The probability in any integrated volume of the domain x, x' depends on Ψ, Ψ^* ; so that when these return to their original values we have the same probability distribution over the domain, *i.e.*, the same sub-state. If they differ from their original values by algebraic factors we still get the same sub-state but with a changed total probability. The same sub-state thus occurs 137 times in a revolution of χ ; or if we count interchanged co-ordinates as the same, it occurs 274 times at intervals of $\pi/137$.

When the same sub-state returns at $\pi/137$, the point (x, x') is not strained to (x', x) but to some other point (α, β) which in general cannot be represented algebraically. For this reason we get no independent determination of the real change of probability due to a displacement $d\chi$. All we learn is that interchange can be effected by a change $\delta\chi = \pi/137$ (involving a real change of ψ) followed by a displacement from (α, β) to (x', x) . The co-ordinate frame being unaltered during the latter change, the real change of ψ is equal to its apparent change $\exp\{(x' - \alpha)\partial/\partial\alpha + (x - \beta)\partial/\partial\beta\}$. It is only by using the result already found in § 10 that we can infer that the changes of χ and of (x, x') provide respectively $1/137$ and $136/137$ of the whole real change of ψ .

The covariant derivatives (real changes) of ψ and Ψ are the same, since these functions differ only through the transformation of the system of reference. Thus it does not matter whether we write ψ or Ψ in the wave-equation written in terms of covariant derivatives. We have introduced Ψ for the purpose of finding the recurrence of sub-states; it will not be required again.

It will be noticed that we have included the independent variable in the volume element. Ordinarily a finite probability is obtained by integration over all co-ordinates other than the independent variable; so that the quantity here considered is related to ordinary probability in the same way that action is to energy. We may call it the "combining probability" of the system; it is the only probability of physical importance, since it is only as it combines with the rest of the universe that physical manifestations of the system occur. When two systems are combined to form one system, their independent variables s_1, s_2 are replaced by $\frac{1}{2}(s_1 + s_2)$ forming the new independent variable for the whole system, and $\frac{1}{2}(s_1 - s_2)$ forming a new internal co-ordinate for the whole system; so that the systems taken together have one extra relation as compared with two disconnected systems. When two separate charges are combined to form a system of two charges this additional relation is the distance between them.

Thus, the throwing out of the independent variable from the probability elements does not occur until the combination of systems is complete, *i.e.*,

ideally until the system comprises the whole universe. Until that stage is reached, it is the combining probability that determines the state. This point is of some importance, for the independent variable must be one of the symmetrical components x_μ , and its omission from the volume element would change the factor from 137 to 136.

§ 12. *Energy of Interchange.*

We deal first with the complication that arises because the angle χ gives a multiple representation of sub-states. To obtain a single representation, let $\gamma = 137\chi$; then the values of γ from 0 to 2π give all the sub-states comprised in the state once only. The corresponding covariant derivative is

$$M_\gamma = M_\chi/137 = (2n + 1) i/137. \quad (12.1)$$

The angular momentum of the state is M_γ , not M_χ . This is clear if we remember that angular and linear momenta belong to the state or wave function and are not in general connected with the displacements of particles. In so far as "motion of charges" is represented in wave mechanics, it is represented by the vector j_μ (sometimes called the charge-current vector) not by the momentum vector p_μ . (In the elementary equation for one charge the two vectors coincide.) When as here the state contains a cyclic co-ordinate γ , we can "ignore" it as in dynamics, inserting the corresponding momentum in the Hamiltonian. In §§ 9, 10 we found a special property of the transformation $\gamma = 137\pi$, which we could turn to account in deducing the magnitude of the covariant derivative; but that must not be confused with the property of γ as the cyclic angle of the state.

It is less easy to recognise this because the other momenta $\partial/\partial x_\mu$ in the wave equation are conjugate to linear displacements, which at first sight seem to have nothing to do with changes of state and to be definable only by reference to the conception of motion of charges. This is because we have neglected the curvature of space-time and so have been forced to adopt a non-relativistic conception of displacement in space-time. In flat space-time there can be no unique scale relation between linear and angular displacements and the question of adopting M_γ rather than M_χ would be meaningless; in fact the wave equation that we are investigating could not exist. In relativity theory dx is defined, not by reference to an anterior conception of linear displacement, but by the corresponding angle of rotation of space-time about its centre of curvature; this definition is contained in the law of gravitation $G_{\mu\nu} = \lambda g_{\mu\nu}$. The rotation is defined as a transformation of state in § 4; in this case it is the

transformation $e^{\frac{1}{2}E_5 d\theta_1}$, where E_5 is the direction towards the centre of curvature and $d\theta_1$ is dx_1 in angular measure. Thus dx_1 like $d\gamma$ is the measure of a transformation of state. Single-valued representation means that when $\int d\theta_1 = 2\pi$ the same state returns; we therefore represent it geometrically by a point indistinguishable from the point of departure, and the point is said to have "gone round the world." In the wave equation $\partial/\partial\theta_1$ (contained in $\partial/\partial x_1$) occurs with a coefficient enormously larger than M_γ ; this, however, is because the term is an averaged sum of momenta corresponding to the pairing of the particle considered with every other particle in the universe (see Paper V).

Thus it will be seen that single-valued representation is implied in all our geometrical comparisons of scale value in different directions.

By (10.5) an angular displacement $d\gamma$ gives a vector displacement

$$r_{12} \cdot d\gamma \cdot \frac{1}{2}P (E_r - F_r). \quad (12.21)$$

Also by (10.4) and (10.6) a linear displacement $dr_{12} = \Sigma l_\mu (dx_\mu - dx'_\mu)$ is represented vectorially by

$$dr_{12} \cdot \frac{1}{2} (E_r - F_r). \quad (12.21)$$

Since the eigenvalue of P is i and P anticommutes with $\frac{1}{2} (E_r - F_r)$, this comparison shows that dr_{12} and $r_{12}d\gamma$ are perpendicular linear elements. That is to say r_{12} is the radius for converting γ to linear scale. (This, for example, decides the question whether r_{12} or $\frac{1}{2}r_{12}$ is the proper radius to take.)

Hence the linear derivative is $\partial/r_{12}\partial\gamma$ and the covariant derivative or linear momentum is

$$\frac{M_\gamma}{r_{12}} = \frac{(2n+1)i}{137r_{12}} \quad (12.3)$$

by (12.1).

The direction of the displacement is, as we have seen, $P \cdot \frac{1}{2} (E_r - F_r)$; but we must distinguish between the symbolic coefficient of the displacement and of the strain; otherwise we lose the i part of the strain ($P + i$). In general the symbolic coefficient of (12.3) would be $(P + i) \cdot \frac{1}{2} (E_r - F_r)$. The most convenient way of describing the direction of the above momentum is as follows. If we choose a representation such that the vector momentum† in the r_{12} direction is *algebraic* the interchange energy-momentum is

$$(P + i) \frac{(2n+1)i}{137r_{12}} \quad (12.4)$$

† By the vector momentum I mean the momentum including its symbolic coefficient.

The correspondence with other momenta in the wave equation is seen if we notice that to $\frac{1}{2}(E_1 + F_1)$, $\frac{1}{2}(E_1 - F_1)$, etc., there should correspond terms in $\frac{1}{2}(P + i)$, $\frac{1}{2}(P - i)$. The latter does not appear because there is no angular momentum associated with it. The factor $\frac{1}{2}$ in the former has been absorbed by using the *diameter* r_{12} of the circle of rotation instead of $\frac{1}{2}r_{12}$.

The second term in (12.4) gives $-(2n + 1)/137r_{12}$. We have required the representation to be such that r_{12} momentum is algebraic; that is to say, the wave equation contains $\partial/\partial r_{12}$ without symbolic coefficient. Hence, as in § 6, the term has the effect of replacing ψ by $r^{(2n+1)/137}\psi$. This has no effect on the eigenvalues of the solution; and indeed the wave function is often left ambiguous as regards an external power of r .† I think that actually this second term is of considerable interest in connection with the packing fraction and other développements, but inasmuch as it does not affect the ordinary theory of energy-levels, etc., it is not counted as an "energy" from the ordinary standpoint.

Thus the interaction energy as ordinarily understood is

$$(2n + 1) \frac{iP}{137r_{12}} = \frac{2n + 1}{137r_{12}} \cdot \frac{1}{4} \sum E_\mu F_\mu. \quad (12.5)$$

§ 13. *Remarks.*

By (12.5) the eigenvalue of the minimum energy of two interchangeable charges is $\pm 1/137r_{12}$. This is accordingly the interchange energy of two electrons or protons or an electron and proton. It is a perhaps unforeseen result that at this stage of the theory protons and electrons are indistinguishable charges; this is in accordance with my theory of protons and electrons (Paper VI) by which the apparent distinction arises through the introduction of curvature of space-time into the wave equation; the dual solution is due to the existence of a quadratic term which vanishes in the flat space-time here treated.

The possibility of values of n other than 0 or -1 in (12.5) suggests that there may be a kind of "excitation" equivalent to a three-fold, five-fold, etc., increase of the electrostatic energy of two charges. I cannot say without further development of the theory whether this is likely to be a practical possibility; it may be prohibited when we deal with systems of more than two charges.

† Cf. Dirac, "Quantum Mechanics," p. 142.

The state can include other cyclic co-ordinates besides γ . Such a co-ordinate must be in a direction which commutes with the $(P + i)$ rotation†; any $\gamma_{\mu\nu}$ rotation satisfies this. If there are several such co-ordinates they must commute with one another, so that the number is in any case very restricted. These additional rotations or strains do not interchange the charges, but if continued through 2π lead back to the original state. There is no earlier return of the original state when account is taken of the variation of the volume element; for the transformation of the volume element (which is highly complicated owing to non-commutation) is not in the same direction as the transformation of ψ and does not combine with it. The transformation $e^{iP \times \psi}$, $e^{iP \times dv}$ is an exceptional property of the symbol P ; for other directions a different symbol occurs in the transformation of ψ and dv , so that the product does not become algebraic until the angle is 2π . Accordingly, the energy of such cyclic co-ordinates is $\pm n/r_{12}$; this is called *energy of excitation*. It appears in the wave equation as in (6.4).

It is to be noticed that $n=0$ is a possible value for other cyclic momenta, but not for the χ momentum which is restricted to odd values of n . That is to say, excitation energy is permissive but interchange energy is compulsory. This is because excitation energy has only to satisfy conditions at a complete revolution, but interchange energy has also to satisfy conditions at the half revolution where the interchange occurs.

The symbol iP can be resolved into two factors

$$\left. \begin{aligned} iP_1 &= \frac{1}{2}(-1 + E_{23}F_{23} + E_{31}F_{31} + E_{12}F_{12}) \\ iP_2 &= \frac{1}{2}(-1 + E_4F_4 + E_5F_5 + E_{45}F_{45}) \end{aligned} \right\}. \quad (13.1)$$

It is easily verified that $-iP_1P_2 = -\frac{1}{2}i \sum E_\mu F_\mu = P$, and that $P_1^2 = -1$, $P_2^2 = -1$ and P_1 commutes with P_2 .

We can always replace the E , F symbols in an equation by any other group of symbols which possess the same mutual commutative properties and the same eigenvalues; so that when the general equation is simplified by a special choice of axes or adaptation to a restricted problem, it may sometimes be possible to use a simpler set of symbols.

† Non-commuting co-ordinates can be used for specifying position in an infinitesimal domain; but to include, for example, a whole rotation from $\theta = 0$ to 2π as required in quantisation co-ordinates must be in commuting (antiperpendicular) directions. The apparent exception afforded by the ordinary space-co-ordinates in four perpendicular directions is due to neglect of space-time curvature. In a curved world rectangular co-ordinates are restricted to an infinitesimal domain.

In particular if for special problems the wave equation is found to contain only the symbols E_{23} , E_{31} , E_{12} , F_{23} , F_{31} , F_{12} , and P , we can set P_2 equal to its eigenvalue i since it commutes with every other symbol in the equation. This will make the equation non-relativistic, for a Lorentz transformation would necessarily re-introduce E_4 , F_4 which do not commute with P_2 . In these conditions the interaction energy (12.5) reduces to

$$\pm iP_1/137r_{12} = \pm \frac{1}{2} (-1 + E_{23}F_{23} + E_{31}F_{31} + E_{12}F_{12})/137r_{12}.$$

A further simplification is possible. To represent the complete set of symbols it is necessary to use four-rowed matrices, but when only the above symbols occur their commutative properties can be represented by Pauli's two-rowed matrices $i\sigma_{12}$, $i\sigma_{23}$, $i\sigma_{31}$. The interaction energy accordingly becomes

$$\pm \frac{1}{2} \frac{(1 + \sigma_{23}\sigma'_{23} + \sigma_{31}\sigma'_{31} + \sigma_{12}\sigma'_{12})}{137r_{12}}. \quad (13.2)$$

Apart from the coefficient 137 determined by the present theory, this is a familiar expression for the energy in non-relativistic theory. Its validity depends on whether it is possible legitimately to get rid of terms containing E_4 , E_5 , etc., in the wave equation by special choice of axes.

So far as I can see, the present deduction of the interchange energy is rigorous and the numerical coefficient is definitely the integer 137. According to current theory the coefficient is $\hbar c/2\pi e^2$. It would follow that $\hbar c/2\pi e^2$ is exactly 137. In stating this result I can only take responsibility for my own part of the deduction; it remains to be seen how far the current identification of the coefficient with $\hbar c/2\pi e^2$ is rigorous. If the neutron is accepted as established observationally the current theory *cannot* be exact; for it gives a wave equation for a proton and electron which has no solutions except those identified with the hydrogen atom. According to my own theory there is an additional quadratic term in the wave equation arising from the curvature of space-time. For a single charge the resulting quadratic equation has two sets of solutions corresponding to electrons and protons; presumably for two charges there are also two sets of solutions corresponding to hydrogen atoms and neutrons, and generally for more complicated systems the two solutions correspond respectively to extra-nuclear and nuclear binding.

Summary.

This paper contains an improved development of the author's theory of the fine-structure constant $\hbar c/2\pi e^2$. It is believed that the value 137 is here

obtained by pure deduction, employing only hypotheses already accepted as fundamental in wave mechanics. Starting from the principle that the wave equation expresses the conservation of probability in a steady state, it is pointed out that, owing to interchangeability, the conservation of probability for two charges is not a simple extension to eight dimensions of the result for one charge in four dimensions. It is necessary to add a new direction of flow of probability between the two configurations corresponding to interchange. The Fermi-Dirac law of antisymmetry is deduced, and it is shown that this quantises the flow in the new direction. It is found that the same state (*i.e.*, distribution of probability) repeats itself 137 times as frequently along the cyclic co-ordinate corresponding to interchange compared with the other cyclic co-ordinates corresponding to excitation. The matrix associated with the interchange energy differs from that given in the earlier papers, and in special cases it reduces to the form $\frac{1}{2}(1 + (\sigma, \sigma'))$ commonly employed in non-relativity theory.

The Viscosity of a Fluid Containing Small Drops of Another Fluid.

By G. I. TAYLOR, F.R.S.

(Received June 30, 1932.)

The viscosity of a fluid in which small solid spheres are suspended has been studied by Einstein as a problem in theoretical hydrodynamics.* Einstein's paper gave rise to many experimental researches on the viscosity of fluids containing solid particles, and it soon became clear that though complete agreement with the theory might be expected when the particles are true spheres, some modification is necessary when the particles are flattened or elongated. The theory of such systems was developed by G. B. Jeffery,† who calculated the motion of ellipsoidal particles in a viscous fluid and their effect on the mean viscosity. Some of his conclusions have been verified by observation.‡

* 'Ann. Physik,' vol. 19, p. 289 (1906), and a correction to that paper, 'Ann. Physik,' vol. 34, p. 591 (1911).

† 'Proc. Roy. Soc.,' A, vol. 102, p. 161 (1922).

‡ G. I. Taylor, 'Proc. Roy. Soc.,' A, vol. 103, p. 58 (1923).

So far no one seems to have extended Einstein's work to liquids containing small drops of another liquid in suspension. The difficulties in the way of a complete theory when solid particles are replaced by fluid drops are almost insuperable, partly because the correct boundary conditions are not known, and partly because a fluid drop would deform under the combined action of viscous forces and surface tension. Even if the boundary conditions were known to be those commonly used in hydrodynamical theory, the calculation of the shape of the deformed drop would be exceedingly difficult. When the radius of the suspended drops or the velocity of distortion of the fluid are small, surface tension may be expected to keep them nearly spherical, and in that case Einstein's analysis may be extended so as to include the case of liquid drops. For this purpose the following assumptions will be made:—

- (1) The drops are so small that they remain nearly spherical.
- (2) There is no slipping at the surface of the drop.
- (3) The tangential stress parallel to the surface is continuous at the surface of the drop, so that any film which may exist between the two liquids merely transmits tangential stress from one fluid to the other.

A general method for analysing the slow motion of viscous fluids has been given by Lamb,* according to which the components of velocity may be regarded as containing three types of terms. The complete expression for one component of velocity is

$$u = \left[\frac{1}{\mu} \sum \frac{r^2}{2(2n+1)} \frac{\partial p_n}{\partial x} + \frac{nr^{2n+3}}{(n+1)(2n+1)(2n+3)} \frac{\partial}{\partial u} \left(\frac{p_n}{r^{2n+1}} \right) \right] \\ + \left[\sum \frac{\partial \phi_n}{\partial x} \right] + \left[\sum \left(z \frac{\partial \chi_n}{\partial y} - y \frac{\partial \chi_n}{\partial z} \right) \right] \quad (1)$$

The terms in the first square bracket are connected with the pressure distribution, which is represented by $p = \sum p_n$, p_n being a solid harmonic function of degree n . The terms in the second bracket represent an irrotational motion which can exist in a field of uniform pressure. The terms in the third bracket represent vortex motion which can exist in a field of uniform pressure. χ_n is an arbitrary function of degree n .

In Einstein's equations the co-ordinate axes were chosen parallel to the principal axes of distortion, and with this choice of axes terms containing χ_n disappear. Einstein used the most general type of flow near a solid sphere,

* "Hydrodynamica," chap. 11.

but little loss of generality is suffered by taking the special case when the mean motion of the whole system is two-dimensional. In this case, if the origin of co-ordinates is taken at the centre of a drop, the flow at great distances from it may be represented by the irrotational flow $\phi_2 = \frac{1}{4} \alpha (x^2 - y^2)$, the constant $\frac{1}{4} \alpha$ being chosen so that the flow is identical, except for a rotation of the whole field, with the familiar case of uniformly shearing laminar flow which is commonly represented by taking the axis of x in the direction of flow when the system is ($u = \alpha y, v = 0$).

Choice of Functions ϕ_n and p_n .

Outside the drop the appropriate functions are

$$\phi_2 = \frac{1}{4} \alpha (x^2 - y^2), \quad \phi_{-3} = B_{-3} a^5 \frac{x^2 - y^2}{r^5}, \quad p_{-3} = \mu A_{-3} a^3 \frac{x^2 - y^2}{r^5} \quad (2)$$

while inside the drop

$$\phi'_2 = B_2 (x^2 - y^2), \quad p_2 = \mu' A_2 a^{-2} (x^2 - y^2) \quad (3)$$

where B_{-3} , A_{-3} , B_2 , A_2 , are constants to be determined by the boundary conditions, μ and μ' are the viscosities of the main body of fluid and the drop, $r^2 = x^2 + y^2 + z^2$, and a is the radius of the drop.

Substituting these expressions in (1), the components of velocity outside the drop are

$$\left. \begin{aligned} u &= \frac{1}{2} A_{-3} a^3 x \frac{x^2 - y^2}{r^5} + B_{-3} a^5 \left[-\frac{5x(x^2 - y^2)}{r^7} + \frac{2x}{r^5} \right] + \frac{1}{2} \alpha x \\ v &= \frac{1}{2} A_{-3} a^3 y \frac{x^2 - y^2}{r^5} + B_{-3} a^5 \left[-\frac{5y(x^2 - y^2)}{r^7} - \frac{2y}{r^5} \right] - \frac{1}{2} \alpha y \\ w &= \frac{1}{2} A_{-3} a^3 z \frac{x^2 - y^2}{r^5} + B_{-3} a^5 \left[-\frac{5z(x^2 - y^2)}{r^7} \right] \end{aligned} \right\} \quad (4)$$

Similarly inside the drop

$$\left. \begin{aligned} u' &= A_2 a^{-2} \left[-\frac{5}{21} x r^2 - \frac{2}{21} x (x^2 - y^2) \right] + 2B_2 x \\ v' &= A_2 a^{-2} \left[-\frac{5}{21} y r^2 - \frac{2}{21} y (x^2 - y^2) \right] - 2B_2 y \\ w' &= A_2 a^{-2} \left[-\frac{2}{21} z (x^2 - y^2) \right] \end{aligned} \right\} \quad (5)$$

Boundary Conditions at $r = a$.

Continuity of velocity requires

$$u = u', \quad v = v', \quad w = w' \quad (6)$$

and the drop remains spherical if

$$ux + vy + wz = 0 \quad (7)$$

At $r = a$,

$$u = \frac{1}{2} A_{-3} a^{-2} x (x^2 - y^2) + B_{-3} a^{-2} [-5x (x^2 - y^2) + 2a^2 x] + \frac{1}{2} \alpha x$$

and

$$u' = \frac{5}{21} A_2 x - \frac{2}{21} A_2 a^{-2} x (x^2 - y^2) + 2B_2 x$$

so that $u = u'$ if

$$\frac{1}{2} A_{-3} - 5B_{-3} = -\frac{2}{21} A_2 \quad (8)$$

and

$$2B_{-3} + \frac{1}{2} \alpha = \frac{5}{21} A_2 + 2B_2; \quad (9)$$

when (8) and (9) are satisfied it will be found that $v = v'$ and $w = w'$, so that all three conditions (6) are satisfied.

To satisfy (7) a formula given by Lamb may be used, namely,

$$xu + yv + zw = \frac{1}{\mu} \sum \frac{nr^2}{2(2n+3)} p_n + \sum u \phi_n$$

This provides a third equation between the four undetermined constants, namely,

$$\frac{1}{2} A_{-3} - 3B_{-3} + \frac{1}{2} \alpha = 0 \quad (10)$$

and the problem will be soluble in this form if only one further equation is necessary in order to satisfy both conditions for continuity of tangential stress.

The components of stress acting across unit area of a spherical surface are p_{rx} , p_{ry} , p_{rz} . The general expression* for p_{rx} is

$$\begin{aligned} rp_{rx} = \sum \left\{ \frac{n-1}{2n+1} r^2 \frac{\partial p_n}{\partial x} + \frac{2n^2 + 4n + 3}{(n+1)(2n+1)(2n+3)} r^{2n+3} \frac{\partial}{\partial x} \left(\frac{p_n}{r^{2n+1}} \right) \right\} \\ + 2\mu \sum (n-1) \frac{\partial \phi_n}{\partial x} + \mu \sum (n-1) \left(y \frac{\partial \chi_n}{\partial z} - z \frac{\partial \chi_n}{\partial y} \right) \end{aligned} \quad (11)$$

In the present case this reduces outside the drop to

$$\frac{rp_{rx}}{\mu} = A_{-3} a^3 \left[\frac{x}{r^3} - \frac{4x}{r^5} (x^2 - y^2) \right] - 8B_{-3} a^5 \left[\frac{2x}{r^5} - \frac{5x}{r^7} (x^2 - y^2) \right] + \alpha x \quad (12)$$

* See Lamb, *l.c.*

and inside the drop to

$$\frac{rp_{rx}}{\mu} = A_2 a^{-2} \left[\frac{16}{21} r^2 x - \frac{19}{21} x (x^2 - y^2) \right] + 4B_2 x, \quad (13)$$

with somewhat similar expressions for p_{ry} and p_{rz} .

Writing

$$\left. \begin{aligned} A_{-3} - 16B_{-3} + \alpha &= \gamma, & \frac{16}{21} A_2 + 4B_2 &= \gamma' \\ 4A_{-3} - 40B_{-3} &= \beta, & \frac{19}{21} A_2 &= \beta' \end{aligned} \right\}, \quad (14)$$

the stress components on the outside of the surface $r = a$ are

$$\left. \begin{aligned} \frac{ap_{rx}}{\mu} &= \gamma x - \beta a^{-2} x (x^2 - y^2) \\ \frac{ap_{ry}}{\mu} &= -\gamma y - \beta a^{-2} y (x^2 - y^2) \\ \frac{ap_{rz}}{\mu} &= -\beta a^{-2} z (x^2 - y^2) \end{aligned} \right\}, \quad (15)$$

while identical expressions serve to express the stress components inside, provided μ, β, γ are replaced by μ', β', γ' . It will be noticed that it is not possible to ensure continuity of all three components of stress. To do so would require both $\mu\beta = \mu'\beta'$ and $\mu\gamma = \mu'\gamma'$, and these equations cannot both be satisfied as well as (8), (9) and (10).

To apply the condition of continuity of tangential stress it is necessary to transform the stress components p_{rx}, p_{ry}, p_{rz} into components p_{rr} acting normal to the surface of the sphere, $p_{r\theta}$ parallel to the surface of the sphere and in a plane passing through the axis of x , $p_{r\phi}$ parallel to the surface of the sphere and perpendicular to the axis of x . The scheme of direction cosines for this transformation is

	p_{rx}	p_{ry}	p_{rz}
p_{rr}	x/r	y/r	z/r
$p_{r\theta}$	$-(1 - x^2/a^2)^{\frac{1}{2}}$	$xy a^{-2} (1 - x^2/a^2)^{\frac{1}{2}}$	$xz a^{-2} (1 - x^2/a^2)^{\frac{1}{2}}$
$p_{r\phi}$	0	$-z a^{-1} (1 - x^2/a^2)^{\frac{1}{2}}$	$y a^{-1} (1 - x^2/a^2)^{\frac{1}{2}}$

Applying this transformation, the stresses at $r = a$ in the outer fluid are

$$\left. \begin{aligned} p_{rr} &= \mu a^{-2} (x^2 - y^2) (\gamma - \beta) \\ p_{r\theta} &= \frac{\mu \gamma x (x^2 - y^2 - a^2)}{a^2 \sqrt{a^2 - x^2}} \\ p_{r\phi} &= \frac{\mu \gamma y z}{a \sqrt{a^2 - x^2}} \end{aligned} \right\}. \quad (16)$$

Identical expressions serve to represent the stress at $r = a$ inside the drop if β, γ, μ are replaced by β', γ', μ' . It will be seen from (16) that continuity of both $p_{r\theta}$ and $p_{r\phi}$ is ensured if

$$\mu \gamma = \mu' \gamma' \text{ or } A_{-3} - 16 B_{-3} + \alpha = \frac{\mu'}{\mu} \left(\frac{16}{21} A_2 + 4 B_2 \right), \quad (17)$$

but when (17) is satisfied p_{rr} is discontinuous at $r = a$.

Determination of A_{-3}, B_{-3}, A_2, B_2 .

The four equations (8), (9), (10), (17) can now be used to determine the constants. The solution is

$$\begin{aligned} A_{-3} &= -\frac{5\alpha}{2} \left(\frac{\mu' + \frac{2}{3}\mu}{\mu' + \mu} \right), & B_{-3} &= -\frac{\alpha}{4} \frac{\mu'}{\mu' + \mu}, & A_2 &= \frac{21\alpha\mu}{4(\mu' + \mu)}, \\ & & & & B_2 &= -\frac{3\alpha\mu}{8(\mu' + \mu)} \end{aligned} \quad (18)$$

Viscosity of the Suspension.

The effect of the presence of solid spheres in suspension on the viscosity of a fluid was shown by Einstein to depend only on p_{-3} , and the same reasoning is still true when the spheres are liquid. In the case of a solid sphere $A_{-3} = -\frac{5\alpha}{2}$; thus it will be seen from (18) that Einstein's expression*

$$\mu^* = \mu (1 + 2.5 \Phi) \quad (19)$$

for the viscosity of a fluid containing solid spheres must be replaced by

$$\mu^* = \mu \left\{ 1 + 2.5 \Phi \left(\frac{\mu' + \frac{2}{3}\mu}{\mu' + \mu} \right) \right\} \quad (20)$$

when the spheres are fluid. In these formulæ μ^* is the mean viscosity and Φ

* 'Ann. Physik,' vol. 34, p. 592 (1911).

is the small proportion of the whole volume occupied by the spheres. The two formulæ are identical, as would be expected when μ' becomes infinite.

The factor $\frac{\mu' + \frac{3}{2}\mu}{\mu' + \mu}$ by which Einstein's term must be multiplied in order to take account of the currents set up inside the drop, may be compared with the factor* $\frac{\mu' + \frac{3}{2}\mu}{\mu' + \mu}$ by which Stokes' expression for the resistance of a solid sphere in a viscous fluid must be multiplied in order to take account of the internal currents in a liquid drop falling through another liquid.

Limits to the Size of Drops contained in a Shearing Fluid.

In order that the drops may be nearly spherical the pressure difference due to viscous forces must be small compared with that due to surface tension, namely, $2T/a$, where T is the surface tension.

The difference in pressure between the inside and outside of the drop is

$$[p_{rr}]_{\text{inside}} - [p_{rr}]_{\text{outside}},$$

and from (16) this is

$$P = [\mu'(\gamma' - \beta') - \mu(\gamma - \beta)] \frac{x^2 - y^2}{a^2},$$

hence substituting from (14) and (18)

$$P = \frac{\alpha\mu}{\mu' + \mu} \left(\frac{19}{4} \mu' + 4\mu \right) \frac{x^2 - y^2}{a^2}. \quad (21)$$

The maximum and minimum values of $\frac{x^2 - y^2}{a^2}$ are ± 1 , so that the drop will be nearly spherical so long as α is small compared with

$$\frac{\mu' + \mu}{\mu \left(\frac{19\mu'}{4} + 4\mu \right)} \left(\frac{2T}{a} \right). \quad (22)$$

On the other hand, the drop whose radius is

$$a = \frac{2T(\mu' + \mu)}{\alpha\mu \left(\frac{19}{4} \mu' + 4\mu \right)}, \quad (23)$$

is of such a size that the disruptive forces due to viscosity tending to burst the drop are about equal to the force due to surface tension which tends to hold it together. An approximate expression of this kind might have some

* Hadamard, 'C.R. Acad. Sci.,' Paris, vol. 152, p. 1735 (1911).

interest in connection with the mechanical formation of emulsions, but since the fluid would certainly be turbulent in any emulsifying machine, the appropriate value to be taken for α would need much consideration.

Summary.

Einstein's expression for the viscosity of a fluid containing solid spheres in suspension is extended so as to include the case when the spheres are liquid. The expression obtained is valid provided the surface tension is great enough to keep the drops nearly spherical. When the rate of distortion of the fluid or the radius of the drop is great enough, the drops tend to break up, and an approximate expression is given for determining the size of the largest drop that can exist in a fluid which is undergoing distortion at any given rate.

On the Theory of Errors and Least Squares.

By HAROLD JEFFREYS, M.A., D.Sc., F.R.S.

(Received July 1, 1932.)

1. In my "Scientific Inference," chapter V, I found that the usual presentation of the theory of errors of observation needed some modification, even where the probability of error is distributed according to the normal law. One change made was in the distribution of the prior probability of the precision constant h . Whereas this is usually taken as uniform (or ignored), I considered it better to assume that the prior probability that the constant lies in a range dh is proportional to dh/h . This is equivalent to assuming that if $h_1/h_2 = h_3/h_4$, h is as likely to lie between h_1 and h_2 as between h_3 and h_4 ; this was thought to be the best way of expressing the condition that there is no previous knowledge of the magnitude of the errors. The relation must break down for very small h , comparable with the reciprocal of the whole length of the scale used, and for large h comparable with the reciprocal of the step of the scale; but for the range of practically admissible values it appeared to be the most plausible distribution.

The argument for this law can now be expressed in an alternative form. The normal law of error is supposed to hold, but the true value x and the precision constant h are unknown. Two measures are made: what is the prob-

ability that the third observation will lie between them? The answer is easily seen to be one-third. For the law says nothing about the order of occurrence of errors of different amounts, and therefore the middle one in magnitude is equally likely to be the first, second, or third made (provided, of course, that we know nothing about the probable range of error already). Now let us see what distribution of prior probability will give this result. We suppose that the prior probability that x and h lie in ranges dx, dh is $f(h) dx dh$. Suppose that the first two observed values are $\pm a$. The probability of obtaining these, given x and h , would be

$$\frac{h}{\sqrt{\pi}} e^{-h^2(x-a)^2} \cdot \frac{h}{\sqrt{\pi}} e^{-h^2(x+a)^2} = \frac{h^2}{\pi} e^{-2h^2(x^2+a^2)}. \quad (1)$$

Hence, if $I(x, h) dx dh$ be the posterior probability that x and h lie in ranges dx, dh ,

$$I(x, h) \propto h^2 f(h) e^{-2h^2(x^2+a^2)}. \quad (2)$$

The probability that the third observation will lie between x_3 and $x_3 + dx_3$ is therefore proportional to

$$\begin{aligned} & \int_0^\infty \int_{-\infty}^\infty I(x, h) \frac{h}{\sqrt{\pi}} e^{-h^2(x_3-x)^2} dh dx \\ & \propto \iint h^3 f(h) \exp[-h^2\{3(x - \tfrac{1}{3}x_3)^2 + 2a^2 + \tfrac{2}{3}x_3^2\}] dx dh \\ & \propto \int_0^\infty h^2 f(h) e^{-2h^2(a^2 + \frac{1}{3}x_3^2)} dh, \end{aligned} \quad (3)$$

some numerical constants being omitted. Then the probability that x_3 will lie between two given values b and c is found by integrating with regard to x_3 between these limits, and is proportional to

$$\int_0^\infty h f(h) e^{-2h^2a^2} \{\operatorname{erf} \sqrt{(\tfrac{2}{3})} hc - \operatorname{erf} \sqrt{(\tfrac{2}{3})} hb\} dh. \quad (4)$$

Our principle is that x_3 is $\frac{1}{3}$ as likely to lie between 0 and a as between 0 and ∞ ; hence

$$\int_0^\infty h f(h) e^{-2h^2a^2} \operatorname{erf} \sqrt{(\tfrac{2}{3})} ha dh = \tfrac{1}{3} \int_0^\infty h f(h) e^{-2h^2a^2} dh. \quad (5)$$

This is to hold for all values of a . If we take ha as a new variable, and $f(h)$ is a power of h , a cancels from the equation; the problem then reduces to determining what power of h we must choose. If, on the other hand, $f(h)$ is

not a power of h , a does not cancel throughout, and (5) cannot hold for all values of a . Hence $f(h)$ is a power of h , and we proceed to investigate whether, if we take it proportional to h^{-1} , (5) will hold. We first take

$$ha = \xi \quad (6)$$

$$\begin{aligned} \int_0^\infty e^{-2\xi^2} \operatorname{erf} \sqrt{\frac{2}{3}} \xi d\xi &= \frac{2}{\sqrt{\pi}} \int_0^\infty e^{-2\xi^2} \left\{ \left(\frac{2}{3}\right)^{\frac{1}{2}} \xi - \frac{1}{3} \left(\frac{2}{3}\right)^{\frac{3}{2}} \xi^3 + \frac{1}{5 \cdot 2!} \left(\frac{2}{3}\right)^{\frac{5}{2}} \xi^5 \dots \right\} d\xi \\ &= \frac{1}{\sqrt{(6\pi)}} \int_0^\infty e^{-u} \left(1 - \frac{1}{3} \frac{u}{3} + \frac{1}{5 \cdot 2!} \left(\frac{u}{3}\right)^2 \dots \right) du \\ &= \frac{1}{\sqrt{(6\pi)}} \left(1 - \frac{1}{3} \cdot \frac{1}{3} + \frac{1}{5} \left(\frac{1}{3}\right)^2 - \frac{1}{7} \left(\frac{1}{3}\right)^3 + \dots \right) \\ &= \frac{1}{\sqrt{(6\pi)}} \sqrt{3} \tan^{-1} \left(\frac{1}{\sqrt{3}} \right) = \frac{1}{6} \left(\frac{1}{2}\pi \right)^{\frac{1}{2}}. \end{aligned} \quad (7)$$

Also

$$\frac{1}{3} \int_0^\infty e^{-2\xi^2} d\xi = \frac{1}{6} \cdot \left(\frac{1}{2}\pi \right)^{\frac{1}{2}}, \quad (8)$$

so that (5) is verified for $f(h) \propto h^{-1}$. This is therefore a possible form for the prior probability. Further, if $hf(h)$ was proportional to h^γ , where $\gamma \neq 0$, we should have to satisfy

$$\int_0^\infty \xi^\gamma e^{-2\xi^2} \{1 - 3 \operatorname{erf} \sqrt{\frac{2}{3}} \xi\} d\xi = 0. \quad (9)$$

If γ is positive we multiply the integrand by a larger factor the greater ξ is, and therefore by larger factors when it is negative than when it is positive; and therefore (9) cannot hold for any positive, or similarly for any negative, value of γ . Thus the only solution is $\gamma = 0$, and

$$f(h) \propto 1/h. \quad (10)$$

2. Let us now reconstruct the theory of Least Squares on the supposition of this distribution of prior probability, as has been done already for the determination of a single quantity from a series of measures. Suppose that we have n linear equations of condition in m unknowns, namely

$$a_{rs}x_s = c_r, \quad (1)$$

where the summation convention is understood and all the c_r are subject to errors following the same normal law. Then for any set of values of x_s and h the prior probability of a given measure of c_r is proportional to

$$\frac{h}{\sqrt{\pi}} e^{-h^2 (a_{rs}x_s - c_r)^2}, \quad (2)$$

and for the complete set to

$$\left(\frac{h}{\sqrt{\pi}}\right)^n e^{-2h^2 W}, \quad (3)$$

where

$$W = \frac{1}{2} \sum_r (a_{rs} x_s - c_r) (a_{rt} x_t - c_r) \quad (4)$$

$$= \frac{1}{2} (b_{st} x_s x_t - 2 d_s x_s + e), \quad (5)$$

with

$$b_{st} = a_{rs} a_{rt}; \quad d_s = a_{rs} c_r; \quad e = c_r^2. \quad (6)$$

The prior probability that h, x_1, x_2, \dots, x_m lie in ranges dh, dx_1, \dots, dx_m is proportional to

$$\frac{1}{h} dh dx_1 \dots dx_m, \quad (7)$$

and hence the posterior probability that they lie in these ranges is proportional to

$$h^{n-1} \exp \{-2h^2 W\} dh dx_1 \dots dx_m. \quad (8)$$

Then the most probable values of the x_t are given by the usual normal equations

$$b_{st} x_t - d_s = 0, \quad (9)$$

and will be denoted by y_t . Put now

$$x_t - y_t = z_t; \quad c_r - a_{rs} y_s = c'_r. \quad (10)$$

Then

$$\begin{aligned} 2W - \sum c_r'^2 &= (a_{rs} x_s - c_r) (a_{rt} x_t - c_r) - (a_{rs} y_s - c_r) (a_{rt} y_t - c_r) \\ &= a_{rs} a_{rt} \{(y_s + z_s) (y_t + z_t) - y_s y_t\} - 2a_{rs} c_r (y_s + z_s) + 2a_{rs} c_r y_s \\ &= b_{st} z_s z_t + 2b_{rt} y_t z_s - 2d_s z_s \\ &= b_{st} z_s z_t, \end{aligned} \quad (11)$$

by (9). Now the c'_r are the residuals when the values y_s are substituted for x_s in the equations of condition. If we write

$$\sum c_r'^2 = c^2, \quad (12)$$

$$2W = c^2 + b_{st} z_s z_t. \quad (13)$$

Also

$$b_{st} z_s z_t = a_{rs} a_{rt} z_s z_t = (a_{rs} z_s)^2, \quad (14)$$

and therefore is a positive form. Then the probability that $z_1 \dots z_m$ and h lie in the given short ranges is proportional to

$$h^{n-1} \exp \{-h^2 (c^2 + b_{st} z_s z_t)\}. \quad (15)$$

We now make the familiar change of variables

$$\left. \begin{aligned} \zeta_1 &= z_2 + \lambda_{2s} z_s & (s = 1, 3, \dots, m) \\ \zeta_2 &= z_3 + \lambda_{3s} z_s & (s = 1, 4, \dots, m) \\ &\dots \dots \dots \\ \zeta_{m-1} &= z_m + \lambda_{m1} z_1 \\ \zeta_m &= z_1 \end{aligned} \right\}. \quad (16)$$

Then $b_{st} z_s z_t$ is reduced to a sum of squares

$$b_1 \zeta_1^2 + b_2 \zeta_2^2 + \dots + b_m \zeta_m^2, \quad (17)$$

when the λ 's are suitably chosen. We require the distribution of the probability of ξ_m for the aggregate of the other variables. The Jacobian of the transformation is 1. Hence the integration with regard to $\zeta_1 \dots \zeta_{m-1}$ gives

$$\begin{aligned} h^{n-1} \exp \{-h^2 (c^2 + b_m z_1^2)\} \int_{-\infty}^{\infty} e^{-h^2 b_1 \zeta_1^2} d\zeta_1 \int_{-\infty}^{\infty} e^{-h^2 b_2 \zeta_2^2} d\zeta_2 \dots \\ \int_{-\infty}^{\infty} e^{-h^2 b_{m-1} \zeta_{m-1}^2} d\zeta_{m-1} = h^{n-1} \left(\frac{\sqrt{\pi}}{h} \right)^{m-1} \frac{1}{(b_1 b_2 \dots b_{m-1})^{\frac{1}{2}}} \exp \{-h^2 (c^2 + b_m z_1^2)\}. \end{aligned} \quad (18)$$

Also
$$b_1 b_2 \dots b_{m-1} b_m = D = ||b_{st}||, \quad (19)$$

and b_1, b_2, \dots, b_{m-1} is the discriminant of the quadratic with z_1 put equal to zero, and therefore is equal to the minor of b_{11} in D . Denoting this minor by B_{11} , we have therefore

$$b_m = D/B_{11}. \quad (20)$$

Now the probability that z_1 lies in a range of dz_1 is, save for a factor independent of z_1 and h ,

$$dz_1 \int_0^\infty h^{n-m} \exp \{-h^2 (c^2 + b_m z_1^2)\} dh = dz_1 \frac{\Pi \left\{ \frac{1}{2} (n - m - 1) \right\}}{2 (c^2 + b_m z_1^2)^{\frac{1}{2} (n - m + 1)}}. \quad (21)$$

To determine the numerical factor we notice that it is certain that z_1 lies between $\pm \infty$; hence finally the probability that z_1 lies in a given dz_1 is

$$\left(\frac{b_m}{\pi} \right)^{\frac{1}{2}} \frac{\Pi \left\{ \frac{1}{2} (n - m - 1) \right\}}{\Pi \left\{ \frac{1}{2} (n - m - 2) \right\}} \frac{1}{c} \left\{ 1 + \frac{b_m z_1^2}{c^2} \right\}^{-\frac{1}{2} (n - m + 1)} dz_1. \quad (22)$$

If there is only one unknown, we have $m = 1$, $c^2 = n\sigma'^2$, where σ' is the

standard deviation. Also $b_m = n$ by inspection of the form taken in this case by $(a_{rs}z_s)^2$. Hence (22) reduces to

$$\frac{\prod \{\frac{1}{2}(n-2)\}}{\prod \{\frac{1}{2}(n-3)\}} \cdot \frac{1}{\sqrt{\pi}} \cdot \frac{1}{\sigma'} \left\{1 + \frac{z^2}{\sigma'^2}\right\}^{-\frac{1}{2}n} dz, \quad (23)$$

which is identical with the form already obtained.*

The posterior probability is therefore not distributed according to the normal law, except as an approximation valid when n is large. For moderate values of n we can estimate the required modification by putting, in the one case,

$$z_1 = cb_m^{-\frac{1}{2}} \tan \theta, \quad (24)$$

and in the other

$$z = \sigma' \tan \theta. \quad (25)$$

In either case the probability of an error between 0 and z_1 or z is

$$\psi(\theta) = \int_0^\theta \cos^{n-m-1} \theta \, d\theta \bigg/ 2 \int_0^{\frac{1}{2}\pi} \cos^{n-m-1} \theta \, d\theta, \quad (26)$$

where in the case of one unknown we put $m = 1$. This function has been tabulated.† The z of the table is the present $\tan \theta$, and the x of the table the present $\sin^2 \theta$. The value of z_1 or z (notation of this paper) corresponding to $\psi = \frac{1}{2}$ is such that the true value is equally likely to lie inside and outside the range from $y_1 - z_1$ to $y_1 + z_1$, and therefore is the probable error as usually defined. $\psi = \frac{1}{2}$ makes the probability of an error numerically greater than z_1 equal to $\frac{1}{2}$; $\psi = \frac{9}{16}$ makes it equal to $\frac{1}{16}$. Using the tables we find for these values of ψ the following values of $u = \tan \theta$.

$n - m + 1.$	$\psi = \frac{1}{2}.$		$\frac{1}{2}.$		$\frac{9}{16}.$	
	u	$u\sqrt{(n-m)}$	u	$u\sqrt{(n-m)}$	u	$u\sqrt{(n-m)}$
2	1.000	1.000	1.732	1.732	6.31	6.31
3	0.577	0.816	0.894	1.265	2.06	2.92
4	0.442	0.766	0.664	1.151	1.359	2.354
5	0.370	0.740	0.550	1.099	1.066	2.132
6	0.325	0.728	0.477	1.066	0.901	2.015
7	0.293	0.718	0.430	1.052	0.793	1.943
10	0.234	0.703	0.341	1.022	0.611	1.833
∞	0	0.674	0	0.967	0	1.644

The values for $n - m$ infinite are derived in the usual way from a table of the error function.

* "Scientific Inference," p. 69.

† Karl Pearson, "Tables for Statisticians and Biometricians," Pt. II, Table XXV; J. Wishart, 'Biometrika,' vol. 17, pp. 68; 469 (1925).

For $n = 2$, $m = 1$, it is exactly as likely as not that $|z|$ is less than σ' , that is, that the true value lies between the observed values. The probable error of the mean is then equal to the standard deviation and to $\sigma'/\sqrt{(n-1)}$.

The usual practice is to suppose that the standard error of the mean is $\sigma'/\sqrt{(n-1)}$ and to compute the probabilities of errors from a table of the error function. The above table shows that for values of n up to 10 the true limits of error are appreciably wider. For each value of ψ an approximate representation was sought in the form

$$u\sqrt{(n-m)} = A \left(1 + \frac{B}{n-m+1} + \frac{C}{(n-m+1)^2} \right), \quad (27)$$

the constants being chosen so as to give agreement at $n = 4, 7, \infty$. The results were, for

$$\begin{array}{lll} \psi = \frac{1}{4} & \dots\dots & A = 0.674, \quad B = 0.34, \quad C = 0.83 \\ \psi = \frac{1}{3} & \dots\dots & A = 0.967, \quad B = 0.41, \quad C = 1.34 \\ \psi = \frac{9}{20} & \dots\dots & A = 1.644, \quad B = 0.66, \quad C = 4.36 \end{array}$$

These were found to give a good representation for $n-m+1 > 4$, and the first two also for $n-m+1 = 3$, but all broke down for $n-m+1 = 2$.

Wishart gives formulæ for $\psi(y)$ in descending powers of n , or $n-m+1$, valid for large values, which can be used if greater accuracy is desired; but the present seem likely to be enough for ordinary purposes.

Where the results of several reductions are to be combined, it may be desirable to have the standard error of each, derived from the formula

$$\sigma_1^2 = \int_{-\infty}^{\infty} E(z_1) z_1^2 dz_1, \quad (28)$$

where $E(z_1)$ is the probability function of (22). This can be written

$$\frac{\sigma_1^2}{\sigma'^2} = \frac{\int_0^{\frac{1}{2}\pi} \cos^{n-2} \theta \tan^2 \theta d\theta}{\int_0^{\frac{1}{2}\pi} \cos^{n-2} \theta d\theta} = \frac{1}{n-3}. \quad (29)$$

when $n > 3$; for $n \leq 3$ it is infinite.

Where the method of least squares is used we have similarly

$$\sigma_1 = c \left(\frac{B_{11}}{(n-m-2)D} \right)^{\frac{1}{2}}. \quad (30)$$

It appears therefore that for small values of $n-m$ the posterior probability falls off so slowly with z that it is not till $n-m = 3$ that it is possible to compute a standard error at all.

Summary.

It is shown that where we have no previous knowledge of the probable degree of precision of a set of observations, the correct procedure is to assume that the probability of a value of h in the range dh is proportional to dh/h .

The result is applied to the theory of least squares, with definite results. It is found that where the number of observations is small, the probability of error is more widely distributed than is given by the usual formula. In the extreme case of two observations and one unknown, the probable error of the mean is equal to the standard deviation.

Contributions to the Mathematical Theory of Epidemics.

II.—*The Problem of Endemicity.*

By W. O. KERMAK and A. G. MCKENDRICK.

(Communicated by Sir Gilbert Walker, F.R.S.—Received April 9, 1932.)

(From the Laboratory of the Royal College of Physicians, Edinburgh.)

Introduction.

In a previous communication* an attempt was made to investigate mathematically the course of an epidemic in a closed population of susceptible individuals. In order to simplify the problem certain definite assumptions were made, namely, that all individuals were equally susceptible, and that death resulted, or complete immunity was conferred, as the result of an attack. The infectivity of the individual and his chances of death or recovery were represented by arbitrary functions, and the chance of a new infection occurring was assumed to be proportional to the product of the infected and susceptible members of the population. In spite of the introduction of the arbitrary functions, it was shown that in general a critical density of population existed, such that if the actual density was less than this, no epidemic could occur, but if it exceeded this by n an epidemic would appear on the introduction of a focus of infection, and further that if n was small relative to the population density, the size of the epidemic would be $2n$ per unit area. It was shown that these

* 'Proc. Roy. Soc.,' A, vol. 115, p. 700 (1927).

conclusions could be readily extended to the case of a metaxenous disease, that is, one in which transmission takes place through an intermediate host.

It is the purpose of the present paper to consider the effect of the continuous introduction of fresh susceptible individuals into the population. It appeared desirable to investigate this point, since it might make it possible to interpret certain aspects of the incidence of disease not only in human communities where there is usually an influx of fresh susceptible individuals either by immigration or by birth, but also in the animal experiments carried out by Topley and others—where fresh animals were introduced at a constant rate into the cages in which cases of disease were already present—from which certain definite results were obtained.

In order to make the present enquiry more general an attempt has also been made to include the effect of the partial immunity which may follow an attack of disease. The susceptibility of the individual who has recovered from an attack is assumed to be a function of the time which has elapsed since his complete recovery. In general this function will either remain constant or will increase with the time. This point is rather important as it will be shown that it enables us to reach certain conclusions even when the function characterising the susceptibility is otherwise quite arbitrary.

It will be seen therefore that in the most general system which is under investigation there will be four classes of individuals: Those which have never experienced an attack of the disease in question, these for convenience we shall call "virgin" (\bar{x}); those who have experienced one or more attacks of the disease but have recovered (x); those who at the moment in question are actually suffering from the disease (y); and, finally, there are the dead (z). All those who enter in to the community first of all pass into the virgin group. In order to make the results as general as possible we shall in the first instance assume that the number so entering this group has a constant component m which may be regarded as representing immigration, and a second component of the form $\bar{\mu}\bar{x} + \mu x + \nu y$, which may be regarded as representing births, the birth-rate of the virgins being $\bar{\mu}$, of the recovered μ , and of the sick ν . By choosing suitable values of m , $\bar{\mu}$, μ and ν the results for various special cases are readily found. Thus taking $\bar{\mu} = \mu = \nu = 0$, we have the case where immigration alone is operative, whilst taking $m = \nu = 0$ and $\bar{\mu} = \mu$, we have the case where there is no immigration, but where all the healthy individuals multiply at the rate μ .

In the mathematical sections which follow, the general case is worked out first of all for constant recovery, death and infectivity rates. Five selected

special cases are then considered, and the results are shown in Table I. The same problem is then worked out for the case in which the recovery, death and infectivity rates are represented by arbitrary general functions instead of by constants. The results obtained are then checked by replacing the arbitrary general functions by the constant coefficients. From the general results the equations for the five selected special cases, this time with arbitrary general functions, are worked out and discussed, and a summary of the findings is given in Table II. Some of the more interesting points which emerge are noted in the general discussion in the final section.

In dealing with the equations referring to constant coefficients the problem of the stability of the steady state is worked out. In the general case with arbitrary general functions, the equations representing the result of the introduction of a small disturbance introduced at a particular time have been formulated, but the detailed discussion of these equations and the stability of the steady state is reserved for a future communication.

The results arrived at in this paper refer to the case where only one disease is operative. It should be emphasised that this implies that the one disease under consideration is thus the only cause of death, and is consequently the only means of balancing in a steady state the increase of population due to immigration and birth. But it is obviously desirable to allow for deaths which may occur from other causes. The methods which are developed in the

Table I.

	$\bar{X}.$	$X.$	$Y.$	$\bar{U}.$	$U.$
General Case	$\frac{d}{k}$	$\frac{l}{k}$	$\frac{m + \bar{\mu}\frac{d}{k} + \mu\frac{l}{k}}{d - v}$	$\frac{d(m + \bar{\mu}\frac{d}{k} + \mu\frac{l}{k})}{d - v}$	$\frac{l(m + \bar{\mu}\frac{d}{k} + \mu\frac{l}{k})}{d - v}$
(1) $m = 0, \bar{\mu} = \mu, v = 0$	$\frac{d}{k}$	$\frac{l}{k}$	$\frac{\mu}{d} \left(\frac{d}{k} + \frac{l}{k} \right)$	$\mu \left(\frac{d}{k} + \frac{l}{k} \right)$	$\frac{l\mu}{d} \left(\frac{d}{k} + \frac{l}{k} \right)$
(2) $m = 0, \bar{\mu} = \mu, v = 0, l = 0$	$\frac{d}{k}$	0	$\frac{\mu}{k}$	$\mu \frac{d}{k}$	0
(3) $\bar{\mu} = \mu = v = 0$	$\frac{d}{k}$	$\frac{l}{k}$	$\frac{m}{d}$	m	$m \frac{l}{d}$
(4) $\bar{\mu} = \mu = v = l = 0$	$\frac{d}{k}$	0	$\frac{m}{d}$	m	0
(5) $\bar{\mu} = \mu = v = d = 0$	0	$\frac{l}{k}$	$n - \frac{l}{k}$	0	0

Table I—(continued).

	V.	\tilde{V} .	\bar{V} .
General Case	$\frac{(d+l)\left(m+\bar{\mu}\frac{d}{\bar{k}}+\mu\frac{l}{\bar{k}}\right)}{d-\nu}$	$\frac{l\left(m+\bar{\mu}\frac{d}{\bar{k}}+\mu\frac{l}{\bar{k}}\right)}{d-\nu}$	$\frac{d\left(m+\bar{\mu}\frac{d}{\bar{k}}+\mu\frac{l}{\bar{k}}\right)}{d-\nu}$
(1) $m=0, \bar{\mu}=\mu, \nu=0$	$\frac{\mu(d+l)}{d}\left(\frac{d}{\bar{k}}+\frac{l}{\bar{k}}\right)$	$\frac{\mu l}{d}\left(\frac{d}{\bar{k}}+\frac{l}{\bar{k}}\right)$	$\mu\left(\frac{d}{\bar{k}}+\frac{l}{\bar{k}}\right)$
(2) $m=0, \bar{\mu}=\mu, \nu=0, l=0$	$\mu\frac{d}{\bar{k}}$	0	$\mu\frac{d}{\bar{k}}$
(3) $\bar{\mu}=\mu=\nu=0$	$\frac{m(d+l)}{d}$	$\frac{ml}{d}$	m
(4) $\bar{\mu}=\mu=\nu=l=0$	m	0	m
(5) $\bar{\mu}=\mu=\nu=d=0$	$l\left(n-\frac{l}{\bar{k}}\right)$	$l\left(n-\frac{l}{\bar{k}}\right)$	0

	W	Equation for α .
General Case	$\frac{d\left(m+\bar{\mu}\frac{d}{\bar{k}}+\mu\frac{l}{\bar{k}}\right)}{d-\nu}$	$\left\{ \alpha^2 + \left\{ \frac{\bar{k}\left(m+\bar{\mu}\frac{d}{\bar{k}}+\mu\frac{l}{\bar{k}}\right)-\bar{\mu}}{d-\nu} \right\} \alpha + \bar{k}\left(m+\bar{\mu}\frac{d}{\bar{k}}+\mu\frac{l}{\bar{k}}\right) = 0 \right.$ $\left. \alpha = -kY \right\}$
(1) $m=0, \bar{\mu}=\mu, \nu=0$	$\mu\left(\frac{d}{\bar{k}}+\frac{l}{\bar{k}}\right)$	$\left\{ \alpha^2 + \frac{\bar{k}\mu l}{d\bar{k}}\alpha + \frac{\bar{k}\mu l}{\bar{k}} + \mu d = 0 \right.$ $\left. \alpha = -kY \right\}$
(2) $m=0, \bar{\mu}=\mu, \nu=0, l=0$	$\mu\frac{d}{\bar{k}}$	$\left\{ \alpha^2 + \mu d = 0 \right.$ $\left. [\alpha = -kY] \right\}$
(3) $\bar{\mu}=\mu=\nu=0$	m	$\left\{ \alpha^2 + \frac{\bar{k}m}{d}\alpha + \bar{k}m = 0 \right.$ $\left. \alpha = -kY \right\}$
(4) $\bar{\mu}=\mu=\nu=l=0$	m	$\left\{ \alpha^2 + \frac{\bar{k}m}{d}\alpha + \bar{k}m = 0 \right.$ $\left. [\alpha = -kY] \right\}$
(5) $\bar{\mu}=\mu=\nu=d=0$	9	$\left\{ \alpha \left\{ \alpha + \bar{k}\left(n-\frac{l}{\bar{k}}\right) \right\} = 0 \right.$ $\left. \alpha = -kY \right\}$

Table II.

	\bar{X} .	X.	Y.	\bar{U} .	U.	Equation for V.	\bar{V} .	\bar{V} .	W.
General Case	$\frac{D}{\bar{K}}$	LVF (V)	NV	DV	LV	$DV = m + \bar{\mu} \frac{D}{\bar{K}} + \mu \text{LVF (V)} + \nu \text{NV}$	LV	DV	DV
(1) $m = 0, \bar{\mu} = \mu, \nu = 0$	$\frac{D}{\bar{K}}$	LVF (V)	NV	DV	LV	$F(V) = \frac{D}{L} \left(\frac{1}{\mu} - \frac{1}{\bar{K}V} \right)$	LV	DV	DV
(2) $m = 0, \bar{\mu} = \mu, \nu = 0$ $l_\theta = 0$	$\frac{1}{\bar{K}}$	0	$\frac{N\mu}{\bar{K}}$	$\frac{\mu}{\bar{K}}$	0	$V = \frac{\mu}{\bar{K}}$	0	$\frac{\mu}{\bar{K}}$	$\frac{\mu}{\bar{K}}$
(3) $\bar{\mu} = \mu = \nu = 0$	$\frac{D}{\bar{K}}$	$L \frac{m}{D} F(V)$	$\frac{mN}{D}$	m	$L \frac{m}{D}$	$V = \frac{m}{D}$	$L \frac{m}{D}$	m	m
(4) $\bar{\mu} = \mu = \nu = 0$ $l_\theta = 0$	$\frac{1}{\bar{K}}$	0	mN	m	0	$V = m$	0	m	m
(5) $m = \bar{\mu} = \mu = \nu = 0$ $d_\theta = 0$	0	VF (V)	NV	0	V	$n = V (F(V) + N)$	V	0	0

following pages can also be applied to the problem as modified by this factor, and it is hoped to deal with this more complicated case later.

1. Constant Rates.

Let us consider a population comprising four classes of individuals :—

- (1) \bar{x} individuals who have never been infected—" virgin " ;
- (2) x partially immune individuals who have recovered from at least one infection ;
- (3) y individuals who are sick, and capable of transmitting the disease ;
- (4) z individuals who have died of the disease.

These numbers will be taken as referring to unit area, so that they really represent population densities. This holds throughout the present paper.

The relationship between the different classes is shown in fig. 1.

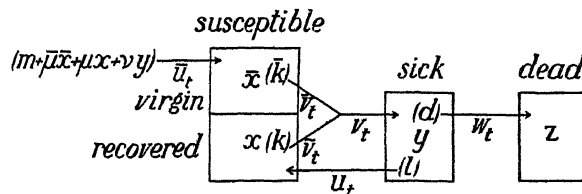


FIG. 1.

In the present section we shall consider the case where the chances of death d , and of recovery l , are independent of the stage of illness, and where the chance

of infection of a particular virgin is proportional to the number of contacts made with individuals already infected, that is to say, it is equal to $\bar{k}y$, whilst the chance of infection of a recovered person is similarly equal to ky . In order to accommodate an immigration of fresh individuals from without at a constant rate, and also reproduction at a constant rate by the three groups—virgin \bar{x} , recovered x , and sick y —the rate at which new individuals accrue is taken to be $\bar{u} = m + \bar{\mu}\bar{x} + \mu x + \nu y$. In particular cases, suitable values of m , $\bar{\mu}$, μ and ν may be taken. Clearly u , the rate at which individuals pass into the recovered group, is equal to ly , and \bar{v} the rate at which virgins turn ill is $\bar{x}\bar{k}y$, because, as explained above, the chance of any particular virgin becoming infected is $\bar{k}y$. Similarly \tilde{v} , the rate at which the recovered turn ill is equal to kxy . Also $v = \bar{v} + \tilde{v}$, measures the incidence of the disease, and $w = dy$ measures the incidence of deaths.

We have then the following equations :—

$$\frac{d\bar{x}}{dt} = \bar{u} - \bar{v} = m + \bar{\mu}\bar{x} + \mu x + \nu y - \bar{k}\bar{x}y, \quad (1)$$

$$\frac{dx}{dt} = u - \tilde{v} = ly - kxy, \quad (2)$$

$$\frac{dy}{dt} = v - w - u = \bar{k}\bar{x}y + kxy - dy - ly. \quad (3)$$

We shall now find the conditions for a steady state, that is such that

$$\frac{d\bar{x}}{dt} = \frac{dx}{dt} = \frac{dy}{dt} = 0.$$

Clearly if \bar{X} , X , Y , U , etc., are the values of \bar{x} , x , y , u , etc., consistent with a steady state, then

$$\left. \begin{aligned} \bar{X} &= \frac{d}{k}, \quad X = \frac{l}{k}, \quad \text{and} \quad Y = \frac{i}{d - \nu} \left(m + \frac{\bar{\mu}d}{k} + \mu \frac{l}{k} \right), \\ \bar{U} &= \bar{V} = W = \frac{d}{d - \nu} \left(m + \frac{\bar{\mu}d}{k} + \mu \frac{l}{k} \right) \\ \text{and} \quad U &= \tilde{V} = \frac{l}{d - \nu} \left(m + \frac{\bar{\mu}d}{k} + \mu \frac{l}{k} \right) \end{aligned} \right\} \quad (4)$$

We shall now consider a system in which the values of \bar{x} , x , y , u , etc., differ only very slightly from \bar{X} , X , Y , U , etc. Such a system may be considered

as the result of an arbitrary disturbance introduced when the system was in a steady state. The behaviour of this system will clearly indicate whether the steady state is stable or unstable.

Let $\bar{x} = \bar{X} + \bar{x}'$, $x = X + x'$, and so on. We substitute these values of \bar{x} , x and y in equations (1), (2) and (3), and we shall neglect any terms in which products of \bar{x}' , x' and y' occur. After reduction we obtain the set of linear differential equations :—

$$\frac{d\bar{x}'}{dt} = \bar{\mu}\bar{x}' + \mu x' + \nu y' - \bar{k}Y\bar{x}' - \bar{k}\bar{X}y', \quad (5)$$

$$\frac{dx'}{dt} = ly' - kXy' - kYx', \quad (6)$$

$$\frac{dy'}{dt} = \bar{k}\bar{X}y' + \bar{k}Y\bar{x}' + kXy' + kYx' - dy' - ly'. \quad (7)$$

Putting $\bar{X} = \frac{d}{k}$, $X = \frac{l}{k}$ as found in (4), these equations reduce to :—

$$\frac{d\bar{x}'}{dt} = \bar{\mu}\bar{x}' + \mu x' + \nu y' - dy' - \bar{k}Y\bar{x}', \quad (8)$$

$$\frac{dx'}{dt} = -kYx', \quad (9)$$

$$\frac{dy'}{dt} = \bar{k}Y\bar{x}' + kYx'. \quad (10)$$

The solution of equation (9) is

$$x' = re^{-kYt}.$$

The values of \bar{x}' and y' will then be given by

$$\bar{x}' = \bar{r}_1 e^{\alpha_1 t} + \bar{r}_2 e^{\alpha_2 t} + \bar{r}_3 e^{-kYt}$$

and

$$y' = s_1 e^{\alpha_1 t} + s_2 e^{\alpha_2 t} + s_3 e^{-kYt},$$

(11)

where α_1 and α_2 are the roots of

$$\begin{vmatrix} \bar{\mu} - \bar{k}Y - \alpha & \nu - d \\ \bar{k}Y & -\alpha \end{vmatrix} = 0. \quad (12)$$

\bar{r}_3 and s_3 are given in terms of r by the equations

$$\left. \begin{aligned} (\bar{\mu} + kY - \bar{k}Y) \bar{r}_3 + \mu r + (\nu - d) s_3 &= 0 \\ \bar{k}Y \bar{r}_3 + kYr + kYs_3 &= 0 \end{aligned} \right\}, \quad (13)$$

and \bar{r}_1 , \bar{r}_2 , s_1 and s_2 are suitably adjusted in accordance with the initial conditions, and with equations (8) and (10). The determinant (12) reduces to

$$\alpha^2 + \left\{ \frac{\bar{k}}{d - \nu} \left(m + \frac{\bar{\mu}d}{\bar{k}} + \mu \frac{l}{k} \right) - \bar{\mu} \right\} \alpha + \bar{k} \left(m + \frac{\bar{\mu}d}{\bar{k}} + \mu \frac{l}{k} \right) = 0. \quad (14)$$

We shall now examine the nature of the roots α_1 and α_2 . Since

$$Y = \frac{1}{d - \nu} \left(m + \frac{\bar{\mu}d}{\bar{k}} + \mu \frac{l}{k} \right)$$

must be positive, a necessary condition for a steady state to exist is $d > \nu$. The coefficient of α in (14) may be written

$$\begin{aligned} & \frac{\bar{k}}{d - \nu} \left(m + \mu \frac{l}{k} \right) + \bar{\mu} \left(\frac{d}{d - \nu} - 1 \right) \\ &= \frac{\bar{k} \left(m + \mu \frac{l}{k} \right) + \bar{\mu} \nu}{d - \nu} \end{aligned}$$

and is therefore positive.

The term independent of α is also positive. The quadratic has therefore either two real negative roots, or two complex roots, the real parts of which are negative. The only exception is when the coefficient of α or the constant term is zero. This can only happen in particular cases which will be considered later. It follows that, in general, the steady states found above correspond to stable states; any disturbance results in either damped vibrations, or in an aperiodic return towards the steady state.

We shall now consider certain special cases, each of which may be worked out separately.

(1) $m = 0$, $\bar{\mu} = \mu$, $\nu = 0$. This means that there is no immigration, and that all healthy persons reproduce at a constant rate.

(2) $m = 0$, $\bar{\mu} = \mu$, $\nu = 0$ and $l = 0$. The conditions are as in case (1), but in addition the disease is fatal. [It is to be noted that in this case the recovered play no part, so that equation (9) disappears, and with it the solution $x = re^{-kYt}$.]

(3) $\bar{\mu} = \mu = \nu = 0$. That is to say, immigration is operative but there is no reproduction.

(4) $\bar{\mu} = \mu = \nu = 0$ and $l = 0$. The conditions are as in case (3) but in addition the disease is fatal. [As in case (2) there is no solution $x = re^{-kYt}$.]

(5) $m = \bar{\mu} = \mu = \nu = 0$ and $d = 0$. That is to say, there are no deaths, births or immigrations. Here we are dealing with a closed population, and with a disease which is never fatal, and may be recovered from. In this case the total population n is fixed and $n = X + Y$ since \bar{X} is zero. It follows that

$$n = \frac{d}{k} + \frac{l}{k} + \frac{1}{d - \nu} \left(m + \frac{\bar{\mu}d}{k} + \mu \frac{l}{k} \right),$$

whence as $m, \bar{\mu}, \mu, \nu$ and d tend to zero

$$\frac{1}{d - \nu} \left(m + \frac{\bar{\mu}d}{k} + \mu \frac{l}{k} \right)$$

must tend to $n - \frac{l}{k}$, that is, to Y .

The possible values of α in this case are there found to be $\alpha = -kY$, $\alpha = -\bar{k}Y$, and $\alpha = 0$. The root $\alpha = 0$ accommodates the circumstance that the disturbance may result in a change in the total number in the closed population by the introduction of new individuals. Of the other two roots, $\alpha = -\bar{k}Y$ appears only when virgins are introduced; if the disturbance consists entirely in changes in the numbers of sick and recovered the only significant root is $\alpha = -kY$. This may readily be verified by treating the case separately.

The results in these special cases are summarised in Table I.

2. Variable Rates.

We shall now consider the more general case when the recovery rate l_θ and the death rate d_θ are no longer independent of the duration of the illness θ . Likewise the chance of a virgin being infected by an infected individual who has been ill for a time θ will now be taken as \bar{k}_θ , whilst a recovered individual

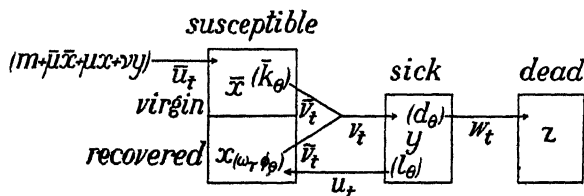


FIG. 2.

who has been recovered for a time τ is assumed to have the chance $k(\tau, \theta)$ of being infected by a person who has been ill for a time θ . We shall assume that $k(\tau, \theta)$ is of the form $\omega_\tau \phi_\theta$, where ω_τ and ϕ_θ are functions of τ and θ respectively. Fig. 2 will make the meanings of these symbols clear. The immigration and the birth-rates remain constant as before. We shall consider that the system is a closed one which has been in existence for an infinite period of time.

The following equations are readily obtained :—

$$\frac{d\bar{x}}{dt} = \bar{u}_t - \bar{v}_t \quad (15)$$

$$\frac{dx}{dt} = u_t - \tilde{v}_t, \quad (16)$$

$$\frac{dy}{dt} = v_t - w_t - u_t, \quad (17)$$

$$\frac{dz}{dt} = w_t. \quad (18)$$

Also

$$\bar{x} = \int_0^\infty \bar{u}_{t\lambda} d\lambda, \quad (19)$$

where $\bar{u}_{t\lambda} d\lambda$ denotes the number of individuals at the time t , who have been born for a period between λ and $\lambda + d\lambda$.

$$x = \int_0^\infty u_{t\tau} d\tau, \quad (20)$$

where $u_{t\tau} d\tau$ denotes the number of susceptibles at a time t who have been recovered for a period between τ and $\tau + d\tau$.

$$y = \int_0^\infty v_{t\theta} d\theta, \quad (21)$$

where $v_{t\theta}$ denotes the number of individuals at a time t who have been sick for a period between θ and $\theta + d\theta$. (Clearly v_{0t} is identical with v_t .)

$$\bar{u}_t = m + \bar{\mu}\bar{x} + \mu x + \nu y, \quad (22)$$

$$u_t = \int_0^\infty l_\theta v_{t\theta} d\theta, \quad (23)$$

$$v_t = \tilde{v}_t + \bar{v}_t, \quad (24)$$

$$\tilde{v}_t = \int_0^\infty \int_0^\infty k(\tau, \theta) v_{t\theta} u_{t\tau} d\theta d\tau, \quad (25)$$

$$\bar{v}_t = \bar{x} \int_0^\infty \bar{k}_\theta v_{t\theta} d\theta, \quad (26)$$

$$w_t = \int_0^\infty d_\theta v_{t\theta} d\theta. \quad (27)$$

It is now necessary to express $\bar{u}_{t\lambda}$, $u_{t\tau}$, etc., in terms of \bar{u}_t , u_t , etc. The rate of infection with time of the group of individuals, the parameters of which, at the moment, lie between t and $t + \delta t$, and λ and $\delta\lambda$, is by definition equal to

$$-\delta t \delta \lambda \int_0^\infty \bar{u}_{t\lambda} \bar{k}_\theta v_{t\theta} d\theta,$$

but it is also given by

$$\delta t \delta \lambda \left(\frac{\partial \bar{u}_{t\lambda}}{\partial t} + \frac{\partial \bar{u}_{t\lambda}}{\partial \lambda} \right) .*$$

Hence we have the equation

$$\begin{aligned} \frac{\partial \bar{u}_{t\lambda}}{\partial t} + \frac{\partial \bar{u}_{t\lambda}}{\partial \lambda} &= - \int_0^\infty \bar{u}_{t\lambda} \bar{k}_\theta v_{t\theta} d\theta, \\ &= - \bar{u}_{t\lambda} \int_0^\infty \bar{k}_\theta v_{t\theta} d\theta, \\ &= - \bar{u}_{t\lambda} \bar{f}(t), \end{aligned} \tag{28}$$

where $\bar{f}(t)$ is written for $\int_0^\infty \bar{k}_\theta v_{t\theta} d\theta$.

Now

$$\frac{\partial v_{t\theta}}{\partial t} + \frac{\partial v_{t\theta}}{\partial \theta} = - (l_\theta + d_\theta) v_{t\theta}$$

whence

$$v_{t\theta} = v_{t-\theta} e^{-\int_0^\theta (l_{\theta'} + d_{\theta'}) d\theta'}, \tag{29}$$

the lower limit of integration being taken as zero, since $v_{t\theta} = v_t$.

Hence

$$\begin{aligned} \bar{f}(t) &= \int_0^\infty \bar{k}_\theta v_{t-\theta} e^{-\int_0^\theta (l_{\theta'} + d_{\theta'}) d\theta'} d\theta, \\ &= \int_0^\infty \bar{k}_\theta v_{t-\theta} N_\theta d\theta, \quad \text{where } N_\theta = e^{-\int_0^\theta (l_{\theta'} + d_{\theta'}) d\theta'} \end{aligned} \tag{30}$$

or

$$\bar{f}(t) = \int_0^\infty \bar{K}_\theta v_{t-\theta} d\theta, \quad \text{where } \bar{K}_\theta \text{ is written for } k_\theta N_\theta. \tag{31}$$

The solution of equation (28) is then

$$\bar{u}_{t\lambda} = \bar{u}_{t-\lambda} e^{-\int_0^\lambda \bar{f}(t-\lambda+\xi) d\xi}.$$

Substituting $\xi' = t - \lambda + \xi$, this may be written in the form

$$\bar{u}_{t\lambda} = \bar{u}_{t-\lambda} e^{-\int_{t-\lambda}^t \bar{f}(\xi') d\xi'}$$

* cf. Kermack and McKendrick (*loc. cit.*), and McKendrick, 'Proc. Edin. Math. Soc.', vol. 44, p. 122 (1926).

which with the substitution $\lambda' = t - \xi'$ becomes

$$\bar{u}_{t\lambda} = \bar{u}_{t-\lambda} e^{-\int_0^\lambda \bar{f}(t-\lambda') d\lambda'} \quad (32)$$

Thus finally equation (19) becomes

$$\bar{x} = \int_0^\infty \bar{u}_{t-\lambda} e^{-\int_0^\lambda \bar{f}(t-\lambda') d\lambda'} d\lambda. \quad (33)$$

Again,

$$\begin{aligned} \frac{\partial u_{t\tau}}{\partial t} + \frac{\partial u_{t\tau}}{\partial \tau} &= - \int_0^\infty k(\theta, \tau) u_{t\tau} v_{t\theta} d\theta, \\ &= - u_{t\tau} \omega_\tau \int_0^\infty \phi_\theta v_{t\theta} d\theta, \end{aligned}$$

since $k(\theta, \tau) = \phi_\theta \omega_\tau$ by hypothesis,

$$\begin{aligned} &= - u_{t\tau} \omega_\tau \int_0^\infty \phi_\theta v_{t-\theta} N_\theta d\theta, \\ &= - u_{t\tau} \omega_\tau \int_0^\infty \Phi_\theta v_{t-\theta} d\theta, \quad \text{where } \Phi_\theta \text{ is written for } \phi_\theta N_\theta \end{aligned} \quad (34)$$

$$= - u_{t\tau} \omega_\tau f(t), \quad (35)$$

where

$$f(t) = \int_0^\infty \Phi_\theta v_{t-\theta} d\theta. \quad (36)$$

Solving equation (35) we have

$$u_{t\tau} = u_{t-\tau} e^{-\int_0^\tau f(t-\tau+\xi) \omega_\xi d\xi}, \quad (37)$$

where it must be remembered that $f(t)$ is a function of $v_{t-\theta}$.

Writing $F(t - \tau, \tau)$ for $e^{-\int_0^\tau f(t-\tau+\xi) \omega_\xi d\xi}$, since it depends on the two variables $t - \tau$ and τ , we have

$$u_{t\tau} = u_{t-\tau} F(t - \tau, \tau). \quad (38)$$

Thus equation (20) becomes

$$x = \int_0^\infty F(t - \tau, \tau) u_{t-\tau} d\tau. \quad (39)$$

Also by equations (21), (29) and (30),

$$y = \int_0^\infty N_\theta v_{t-\theta} d\theta, \quad (40)$$

and by equation (23)

$$\begin{aligned} u_t &= \int_0^\infty l_\theta v_{t-\theta} N_\theta d\theta, \\ &= \int_0^\infty L_\theta v_{t-\theta} d\theta, \quad \text{where } L_\theta \text{ is written for } l_\theta N_\theta. \end{aligned} \quad (41)$$

By equation (25)

$$\begin{aligned} \tilde{v}_t &= \int_0^\infty \phi_\theta v_{t-\theta} d\theta \int_0^\infty \omega_\tau u_{t-\tau} d\tau, \\ &= \int_0^\infty \Phi_\theta v_{t-\theta} d\theta \int_0^\infty \omega_\tau F(t-\tau, \tau) u_{t-\tau} d\tau \quad (\text{by (34) and (38)}), \\ &= \int_0^\infty \Phi_\theta v_{t-\theta} d\theta \int_0^\infty G(t-\tau, \tau) u_{t-\tau} d\tau, \quad \text{where } G(t-\tau, \tau) = \omega_\tau F(t-\tau, \tau). \end{aligned} \quad (42)$$

By equations (26), (29) and (30)

$$\begin{aligned} \bar{v}_t &= \bar{x} \int_0^\infty \bar{k}_\theta v_{t-\theta} N_\theta d\theta, \\ &= \bar{x} \int_0^\infty \bar{K}_\theta v_{t-\theta} d\theta, \quad \text{by (31)} \\ &= \bar{x} \bar{f}(t), \quad \text{by (31)}. \end{aligned} \quad (43)$$

Finally by equation (27)

$$\begin{aligned} w_t &= \int_0^\infty d_\theta v_{t-\theta} N_\theta d\theta, \\ &= \int_0^\infty D_\theta v_{t-\theta} d\theta, \quad \text{where } D_\theta = d_\theta N_\theta. \end{aligned} \quad (44)$$

We shall now find the conditions for a steady state. We shall assume that in this condition

$$\bar{x} = \bar{X}, x = X, y = Y, \bar{u}_t = \bar{U}, u_t = U, v_t = V, \tilde{v}_t = \tilde{V}, \bar{v}_t = \bar{V} \text{ and } w_t = W.$$

We shall also adopt the notation

$$L = \int_0^\infty L_\theta d\theta, \quad D = \int_0^\infty D_\theta d\theta, \quad \bar{K} = \int_0^\infty \bar{K}_\theta d\theta, \quad \text{etc.} \quad (45)$$

We shall in general assume that L_θ , D_θ , \bar{K}_θ , etc., which are, of course, never negative, are zero for all values of θ greater than θ_0 , where θ_0 is some finite

quantity. This ensures the convergency of the integrals denoted by L , D , \bar{K} , etc.

Clearly

$$L + D = \int_0^\infty (l_\theta + d_\theta) e^{-\int_0^\theta (l_{\theta'} + d_{\theta'}) d\theta'} d\theta = 1, \quad (46)$$

since by assumption $L_\theta = D_\theta = 0$, when $\theta = \infty$.

By equations (15) and (22),

$$\bar{U} = \bar{V} = m + \bar{\mu}\bar{X} + \mu X + \nu Y \quad (47)$$

$$U = \tilde{V}, \text{ by equation (16),} \quad (48)$$

$$V = W + U, \text{ by equation (17).} \quad (49)$$

By equation (33)

$$\bar{X} = \bar{U} \int_0^\infty e^{-\int_0^\lambda \tilde{f}(t-\lambda') d\lambda'} d\lambda,$$

but by (31)

$$\begin{aligned} \tilde{f}(t) &= \int_0^\infty \bar{K}_\theta v_{t-\theta} d\theta \\ &= V \int_0^\infty \bar{K}_\theta d\theta \\ &= \bar{K}V, \text{ by (45),} \end{aligned} \quad (50)$$

hence

$$\begin{aligned} \bar{X} &= \bar{U} \int_0^\infty e^{-\bar{K}V\lambda} d\lambda, \\ &= \frac{\bar{U}}{\bar{K}V}. \end{aligned} \quad (51)$$

Similarly by equation (39)

$$X = U \int_0^\infty F(t-\tau, \tau) d\tau,$$

but

$$F(t-\tau, \tau) = e^{-\int_0^\tau f(t-\tau+\xi) \omega_\xi d\xi},$$

and

$$\begin{aligned} f(t) &= \int_0^\infty \Phi_\theta v_{t-\theta} d\theta, \text{ by (36),} \\ &= V\Phi, \end{aligned} \quad (52)$$

hence

$$\int_0^\infty F(t-\tau, \tau) d\tau = \int_0^\infty e^{-V\Phi \int_0^\tau \omega_\xi d\xi} d\tau = F(V). \quad (53)$$

[It is to be noted that Φ is independent of V .]

Thus finally

$$X = UF(V). \quad (54)$$

By equation (40)

$$Y = NV; \quad (55)$$

by equation (41)

$$U = LV, \quad (56)$$

hence

$$X = LVF(V). \quad (57)$$

By equations (42) and (48)

$$\tilde{V} = \Phi VUG(V) = U, \quad (58)$$

since

$$\begin{aligned} G(V) &= \int_0^\infty G(t - \tau, \tau) d\tau \\ &= \int_0^\infty \omega F(t - \tau, \tau) d\tau, \text{ by (42),} \\ &= \int_0^\infty \omega e^{-\nu \Phi \int_0^\tau \omega_\xi d\xi} d\tau, \\ &= \frac{1}{\Phi V}. \end{aligned} \quad (59)$$

Also by equation (26)

$$\bar{V} = \bar{X}\bar{K}V, \quad (60)$$

and by equation (27)

$$W = DV. \quad (61)$$

From these relations it follows that

$$\left. \begin{aligned} U &= \tilde{V} = LV \\ \bar{U} &= \bar{V} = W = DV \\ \bar{X} &= \frac{D}{\bar{K}} \\ X &= LVF(V) \\ Y &= NV \end{aligned} \right\}. \quad (62)$$

and

Hence, making the necessary substitutions in equation (47), viz.,

$$\bar{U} = m + \bar{\mu}\bar{X} + \mu X + \nu Y,$$

we have

$$DV = m + \bar{\mu} \frac{D}{\bar{K}} + \mu LVF(V) + \nu NV, \quad (63)$$

an equation which determines V .

Also the total number of individuals is

$$\begin{aligned} n &= \bar{X} + X + Y \\ &= V \left(\frac{D}{\bar{K}} + LF(V) + N \right). \end{aligned} \quad (64)$$

We shall now consider the nature of the real positive roots of equation (63).

Let

$$\Theta(V) = \mu LF(V) + \nu N - D + \frac{m}{V} + \frac{\bar{\mu}D}{KV}, \quad (65)$$

and let us write $F(V)$ in the form $\int_0^\infty e^{-\Phi V \bar{\Omega}_\tau} d\tau$, where $\bar{\Omega}_\tau$ is written for $\int_0^\tau \omega_\xi d\xi$. In general $\bar{\Omega}_\tau$ will have the following characteristics. It may be zero when $\tau = 0$ and $\tau = \varepsilon$, as ω_ξ may be zero for small values of ε , that is to say, immediately after recovery there may be complete immunity. It will then increase monotonically as ω_ξ is always positive. In general for large values of ξ , say $\xi > \eta$, ω_ξ will become sensibly constant and equal to ω . For large values of τ then,

$$\begin{aligned} \bar{\Omega}_\tau &= \int_0^\tau \omega_\xi d\xi = \int_0^\eta \omega_\xi d\xi + \int_\eta^\tau \omega_\xi d\xi, \\ &= \bar{\Omega}_\eta + \omega(\tau - \eta). \end{aligned} \quad (66)$$

It may easily be shown that for finite positive values of V , $F(V)$ is always finite, and that as V increases $F(V)$ always decreases; also as V increases $\nu N - D + \frac{m}{V} + \frac{\bar{\mu}D}{KV}$ always decreases, so that $\Theta(V)$ always decreases as V increases. Further, for small values of V , $\Theta(V)$ is very large and positive. Therefore there may be either one positive real root, or no positive real roots according as $\Theta(V = \infty)$ is negative or positive.

Now

$$F(V) = \int_0^\varepsilon e^{-\Phi V \bar{\Omega}_\tau} d\tau + \int_\varepsilon^\eta e^{-\Phi V \bar{\Omega}_\tau} d\tau + \int_\eta^\infty e^{-\Phi V \bar{\Omega}_\tau} d\tau. \quad (67)$$

The first integral is simply equal to ε . The second integral is equal to $(\eta - \varepsilon)e^{-\Phi V \bar{\Omega}_\eta}$, where $\bar{\Omega}_\eta$ is some value of $\bar{\Omega}_\tau$ between zero and $\bar{\Omega}_\eta$ and is therefore finite. The third integral becomes $\int_\eta^\infty e^{-\Phi V [\bar{\Omega}_\eta + \omega(\tau - \eta)]} d\tau$, by

equation (66). When V is very great the second integral vanishes since $\bar{\Omega}_\pi$ is finite; and the third integral becomes

$$\begin{aligned} & e^{-\Phi V \bar{\Omega}_\eta + \Phi V \omega_\eta} \times \int_\eta^\infty e^{-\Phi V \omega_\tau} d\tau, \\ &= e^{-\Phi V \bar{\Omega}_\eta + \Omega V \omega_\eta} \times \frac{e^{-\Phi V \omega_\eta}}{\Phi V \omega}, \\ &= \frac{e^{-\Phi V \bar{\Omega}_\eta}}{\Phi V \omega} \end{aligned}$$

which tends to zero as V tends to infinity.

Thus $F(V)_{V \rightarrow \infty} \rightarrow \varepsilon$, and $\Theta(V)$ tends to

$$\mu L \varepsilon + \nu N - D. \quad (68)$$

Consequently $\Theta(V) = 0$ has no real root if $D - \nu N - \mu L \varepsilon$ is negative, and has one real root if this expression is positive. It follows that no steady state exists if $D - \nu N - \mu L \varepsilon$ is negative, and that a unique steady state exists if it is positive. It is to be noted that ε is equal to the time during which immunity is absolute.

Let us now consider the effect of varying the immigration rate m , or the birth-rates $\bar{\mu}$, μ and ν , upon the number turning ill V . Equation (65) may be regarded as an equation giving V in terms of m , $\bar{\mu}$, μ and ν , hence

$$\frac{\partial \Theta}{\partial V} \delta V + \frac{\partial \Theta}{\partial m} \delta m + \frac{\partial \Theta}{\partial \bar{\mu}} \delta \bar{\mu} + \frac{\partial \Theta}{\partial \mu} \delta \mu + \frac{\partial \Theta}{\partial \nu} \delta \nu = 0,$$

or

$$\frac{\partial \Theta}{\partial V} \delta V + \frac{\delta m}{V} + \frac{D}{KV} \delta \bar{\mu} + LF \delta \mu + N \delta \nu = 0.$$

Consequently

$$\frac{\partial V}{\partial m} = -\frac{1}{V\Theta'}, \quad \frac{\partial V}{\partial \bar{\mu}} = -\frac{D}{KV\Theta'}, \quad \frac{\partial V}{\partial \mu} = -\frac{LF}{\Theta'} \text{ and } \frac{\partial V}{\partial \nu} = -\frac{N}{\Theta'}. \quad (69)$$

It has been shown above that Θ' is always negative, hence

$$\frac{\partial V}{\partial m}, \quad \frac{\partial V}{\partial \bar{\mu}}, \quad \frac{\partial V}{\partial \mu} \text{ and } \frac{\partial V}{\partial \nu}$$

are each positive. It follows that an increase of the immigration or birth-rates results in an increase in the number turning ill and *vice versa*.

Further, let us write $T = V/n$, so that $100T$ is the percentage rate of incidence of fresh cases of the disease. Then

$$\begin{aligned}\frac{1}{T} &= \frac{n}{V} = \frac{X}{V} + \frac{\bar{X}}{V} + \frac{Y}{V}, \\ &= LF + \frac{D}{KV} + N.\end{aligned}$$

Hence

$$-\frac{1}{T^2} \frac{\partial T}{\partial V} = L \frac{dF}{dV} - \frac{D}{KV^2}.$$

But

$$F = \int_0^\infty e^{-\Phi V} \int_0^\tau \omega_\xi d\xi d\tau.$$

Therefore

$$\frac{\partial F}{\partial V} = -\Phi \int_0^\infty \int_0^\tau \omega_\xi d\xi e^{-\Phi V} \int_0^\tau \omega_\xi d\xi d\tau = -\Phi I \Gamma,$$

where

$$\Gamma = \int_0^\infty \int_0^\tau \omega_\xi d\xi e^{-\Phi V} \int_0^\tau \omega_\xi d\xi d\tau,$$

which is always positive. Thus

$$\frac{\partial T}{\partial V} = T^2 \left(L \Phi \Gamma + \frac{D}{KV^2} \right). \quad (70)$$

Hence $\frac{\partial T}{\partial V}$ is positive, and as

$$\frac{\partial V}{\partial m}, \frac{\partial V}{\partial \bar{\mu}}, \frac{\partial V}{\partial \mu} \text{ and } \frac{\partial V}{\partial \nu}$$

are each positive, it follows that

$$\frac{\partial T}{\partial m}, \frac{\partial T}{\partial \bar{\mu}}, \frac{\partial T}{\partial \mu} \text{ and } \frac{\partial T}{\partial \nu}$$

are each positive. Since T is proportional to the percentage rate of incidence of fresh cases of the disease, it follows from the above, that if a condition of equilibrium exists, the percentage rate of incidence increases as the immigration or birth-rates increase.

In the previous section dealing with constant rates, the question of the stability of the steady states has been treated. In the more general problem, now under consideration, the question of the stability of the steady states is much more difficult. We shall at present confine ourselves to the formulation

of the equations which describe the behaviour of the system after the introduction of a small disturbance at a particular time $-T$. Considerable progress has been made with the investigation of these equations, and it is hoped to present the results in a later communication. It is meanwhile convenient to formulate the equations here in order to avoid repetition in the future of the details of the epidemiological problem under consideration.

We shall consider, then, a system in which a steady state has existed for all time, but into which, at a time $-T$ a small disturbance was introduced in the form of a number of virgins ${}_0\bar{u}_\lambda d\lambda_1$ of age between λ_1 and $\lambda_1 + d\lambda_1$, together with a number of recovered persons of whom ${}_0u_{\lambda_1}d\lambda_1$ had recovered for a period between λ_1 and $\lambda_1 + d\lambda_1$, and a number of infected individuals of whom ${}_0v_{\theta_1}d\theta_1$ had been infected for a period between θ_1 and $\theta_1 + d\theta_1$.

We shall first of all write down the exact equations which define the system. To prevent confusion we shall, in what follows, limit \bar{u} , u and v so that they refer solely to new virgins, recovered persons, or infected cases occurring as the result of the working of the postulated processes within the system, and do not include individuals introduced from without. The values \bar{x} , x and y , on the other hand, include all the virgins, recovered persons, and infected individuals independently of how they originated.

It follows that the equations (22), (41), (42) and (43) for \bar{u}_t , u_t , and v_t hold without modification, whilst to the equations for \bar{x} , x and y (33), (39) and (40) an extra term must be added to accommodate the introduction of the extra individuals.

Thus

$$y = \int_0^\infty N_\theta v_{t-\theta} d\theta + \int_0^\infty v_{t\theta_1} d\theta_1, \quad (71)$$

where $v_{t\theta_1} d\theta_1$ indicates the number surviving at the time t , of the ${}_0v_{\theta_1} d\theta_1$ introduced at the time $-T$, who at that time were of age between θ_1 and $\theta_1 + d\theta_1$.

Similarly

$$\bar{x} = \int_0^\infty \bar{u}_{t-\lambda} e^{-\int_0^\lambda \bar{f}(t-\lambda') d\lambda'} d\lambda + \int_0^\infty \bar{u}_{t\lambda_1} d\lambda_1, \quad (72)$$

and

$$x = \int_0^\infty F(t-\tau, \tau) u_{t-\tau} d\tau + \int_0^\infty u_{t\tau_1} d\tau_1. \quad (73)$$

In these equations and in those for \bar{u} , u , v , etc., we write

$$\begin{aligned} \bar{x} &= \bar{X} + \bar{x}', & x &= X + x', & y &= Y + y', & \bar{u}_t &= \bar{U} + \bar{u}'_t, & u_t &= U + u'_t, \\ v_t &= V + v'_t, & \bar{v}_t &= \bar{V} + \bar{v}'_t, & \tilde{v}_t &= \tilde{V} + \tilde{v}'_t, & \text{and} & w_t &= W + w'_t, \end{aligned}$$

where \bar{X} , X , Y , \bar{U} , etc., are the steady state values found above, and \bar{x}' , x' , y' , etc., are zero for all values of $t < -T$. The values \bar{x}' , x' , y' , etc., are very small quantities, provided that

$$\int_0^\infty \bar{u}_{t\lambda_1} d\lambda_1, \quad \int_0^\infty u_{t\tau_1} d\tau_1, \quad \text{and} \quad \int_0^\infty v_{t\theta_1} d\theta_1$$

are small, for they necessarily tend to zero as the disturbance represented by these integrals tends to zero.

After simple reductions, and, where necessary, neglecting small quantities of higher order than the first, we have

$$\frac{d\bar{x}'}{dt} = \bar{u}'_t - \bar{v}'_t, \quad (74)$$

$$\frac{d'x}{dt} = u'_t - \bar{v}'_t, \quad (75)$$

$$\frac{dy'}{dt} = v'_t - w'_t - u'_t, \quad (76)$$

$$\bar{u}'_t = \bar{u}\bar{x}' + \mu x' + \nu y', \quad (77)$$

$$u'_t = \int_0^\beta L_\theta v'_{t-\theta} d\theta, \quad (\beta = \infty) \quad (78)$$

$$v'_t = \bar{v}' + \bar{v}', \quad (79)$$

$$\bar{v}'_t = \bar{x}'KV + \bar{X} \int_0^\beta \bar{K}_\theta v'_{t-\theta} d\theta, \text{ from (43) } (\beta = \infty) \quad (80)$$

$$w'_t = \int_0^\beta D_\theta v'_{t-\theta} d\theta. \quad (81)$$

Also from (71)

$$y'_t = \int_0^\beta N_\theta v'_{t-\theta} d\theta + \int_0^\infty v_{t\theta} d\theta_1 \quad (\beta = \infty). \quad (82)$$

The following reductions are more complicated.

By equation (72) we have

$$\bar{x} = \bar{X} + \bar{x}' = \int_0^\beta (\bar{U} + \bar{u}'_{t-\lambda}) e^{-\int_0^\lambda f(t-\lambda') d\lambda'} d\lambda + \int_0^\infty \bar{u}_{t\lambda_1} d\lambda_1, \quad (\beta = \infty),$$

but by (31)

$$\begin{aligned} \bar{f}(t - \lambda') &= \int_0^\infty \bar{K}_\theta v_{t-\lambda'-\theta} d\theta, \\ &= \int_0^\infty \bar{K}_\theta (V + v'_{t-\lambda'-\theta}) d\theta, \\ &= \bar{K}V + \int_0^\infty \bar{K}_\theta v'_{t-\lambda'-\theta} d\theta. \end{aligned} \quad (83)$$

Hence

$$\begin{aligned}\bar{X} + \bar{x}' &= \int_0^\infty (\bar{U} + \bar{u}'_{t-\lambda}) e^{-\bar{K}V\lambda} e^{-\int_0^\lambda \int_0^\infty \bar{K}_\theta v'_{t-\lambda'-\theta} d\theta d\lambda'} d\lambda + \int_0^\infty \bar{u}_{t\lambda_1} d\lambda_1, \\ &= \int_0^\infty (\bar{U} + \bar{u}'_{t-\lambda}) e^{-\bar{K}V\lambda} \left(1 - \int_0^\lambda \int_0^\infty \bar{K}_\theta v'_{t-\lambda'-\theta} d\theta d\lambda'\right) d\lambda + \int_0^\infty \bar{u}_{t\lambda_1} d\lambda_1, \\ &= \frac{\bar{U}}{\bar{K}V} + \int_0^\infty \bar{u}'_{t-\lambda} e^{-\bar{K}V\lambda} d\lambda - \bar{U} \int_0^\infty e^{-\bar{K}V\lambda} \int_0^\lambda \int_0^\infty \bar{K}_\theta v'_{t-\lambda'-\theta} d\theta d\lambda' d\lambda \\ &\quad + \int_0^\infty \bar{u}_{t\lambda_1} d\lambda_1,\end{aligned}$$

whence

$$\bar{x}' = \int_0^\beta \bar{u}'_{t-\lambda} e^{-\bar{K}V\lambda} d\lambda - \bar{U} \int_0^\beta e^{-\bar{K}V\lambda} \int_0^\lambda \int_0^{\beta-\lambda'} \bar{K}_\theta v'_{t-\lambda'-\theta} d\theta d\lambda' d\lambda + \int_0^\infty \bar{u}_{t\lambda_1} d\lambda_1, \quad (84)$$

where $\beta = \infty$.

Also by equation (42)

$$\bar{v}_t = \int_0^\infty \Phi_\theta (V + v'_{t-\theta}) d\theta \int_0^\infty G(t - \tau, \tau) (U + u'_{t-\tau}) d\tau.$$

Writing

$$\begin{aligned}G(t - \tau, \tau) &= \omega_\tau F(t - \tau, \tau) \\ &= \omega_\tau F_\tau + \omega_\tau F'(t - \tau, \tau),\end{aligned}$$

where $F'(t - \tau, \tau)$ contains only small quantities of the first order, small quantities of the second and higher orders being neglected, we have

$$\begin{aligned}\tilde{V} + \bar{v}' &= \int_0^\infty \Phi_\theta (V + v'_{t-\theta}) d\theta \int_0^\infty (U + u'_{t-\tau}) \{\omega_\tau F_\tau + \omega_\tau F'(t - \tau, \tau)\} d\tau \\ &= \left[\Phi V + \int_0^\infty \Phi_\theta v'_{t-\theta} d\theta \right] \int_0^\infty \{U \omega_\tau F_\tau + U \omega_\tau F'(t - \tau, \tau) + \omega_\tau F_\tau u'_{t-\tau}\} d\tau.\end{aligned}$$

Thus writing $G = \int_0^\infty \omega_\tau F_\tau d\tau$, and $G' = \int_0^\infty \omega_\tau F'(t - \tau, \tau) d\tau$, so that

$G_t = G + G'_t$ to a first order of approximation; since $\tilde{V} = \Phi V U G$,

$$\bar{v}' = \Phi V U \int_0^\beta \omega_\tau F'(t - \tau, \tau) d\tau + \Phi V \int_0^\beta \omega_\tau F_\tau d\tau + U G \int_0^\beta \Phi_\theta v'_{t-\theta} d\theta. \quad (85)$$

Again by equation (73)

$$\begin{aligned}X + x' &= \int_0^\infty (U + u'_{t-\tau}) \{F_\tau + F'(t - \tau, \tau)\} d\tau + \int_0^\infty u_{t\tau_1} d\tau_1, \\ &= \int_0^\infty U F_\tau d\tau + \int_0^\infty U F'(t - \tau, \tau) d\tau + \int_0^\infty F_\tau u'_{t-\tau} d\tau + \int_0^\infty u_{t\tau_1} d\tau_1, \\ &= U F + U \int_0^\infty F'(t - \tau, \tau) d\tau + \int_0^\infty F_\tau u'_{t-\tau} d\tau + \int_0^\infty u_{t\tau_1} d\tau_1.\end{aligned}$$

Therefore since $X = UF$

$$x' = U \int_0^\beta F'(t - \tau, \tau) d\tau + \int_0^\beta F_t u'_{t-\tau} d\tau + \int_0^\infty u_{t\tau_1} d\tau_1 \quad (\beta = \infty). \quad (86)$$

Thus \bar{u}' , \bar{u}' , v' , \bar{x}' , x' , etc., are determined by equations (77) to (86). As, however, all these functions are zero when $t < -T$, it is easily seen that the upper limit β of the integrals involving these unknown functions may be put equal to $t + T$ instead of infinity. It follows that these functions at any point, which have now been defined in terms of finite integrals of their values up to that point, must remain finite and single valued for all finite values of t , provided that the initial disturbance is defined unambiguously.

It is now necessary to express $v_{t\theta_1}$ in terms of ${}_0v_{\theta_1}$ the number actually introduced at the time $-T$. It can be shown that

$$v_{t\theta_1} = {}_0v_{\theta_1} e^{-\int_{\theta_1}^{t+T+\theta_1} (l_{\theta'} + d_{\theta'}) d\theta'}. \quad (87)$$

Also

$$\bar{u}_{t\lambda_1} = {}_0\bar{u}_{\lambda_1} e^{-\int_{\lambda_1}^{t+T+\lambda_1} \bar{f}(t-\lambda') d\lambda'}. \quad (88)$$

Since $\bar{f}(t) = \int_0^\infty \bar{K}_\theta v_{t-\theta} d\theta$, therefore to a first approximation,

$$\bar{f}(t) = \int_0^\infty \bar{K}_\theta V d\theta = \bar{K}V,$$

hence

$$\bar{u}_{t\lambda_1} = {}_0\bar{u}_{\lambda_1} e^{-\bar{K}V(t+T)}. \quad (89)$$

It may also be shown that

$$u_{t\tau_1} = {}_0u_{\tau_1} \frac{F(-T - \tau_1, t + T + \tau_1)}{F(-T - \tau_1, \tau_1)}. \quad (90)$$

But $F(t - \tau, \tau) = e^{-\int_0^\tau f(t-\tau+\xi) \omega_\xi d\xi}$, and to a first approximation $f(t) = \Phi V$, therefore $F(t - \tau, \tau) = e^{-\Phi V}$, so that

$$\frac{F(-T - \tau_1, t + T + \tau_1)}{F(-T - \tau_1, \tau_1)} = e^{-\Phi V \int_{\tau_1}^{t+T+\tau_1} \omega_\xi d\xi}. \quad (91)$$

From the above we obtain finally the five equations:—

$$\begin{aligned} \bar{x}' = & \int_0^{t+T} \bar{u}'_{t-\lambda} e^{-\bar{K}V\lambda} d\lambda - \bar{U} \int_0^{t+T} e^{-\bar{K}V\lambda} \int_0^\lambda \int_0^{t+T-\lambda'} \bar{K}_\theta v'_{t-\lambda'-\theta} d\theta d\lambda' d\lambda \\ & + \int_0^\infty \bar{u}_{t\lambda_1} d\lambda_1, \end{aligned} \quad (92)$$

$$x' = U \int_0^{t+T} F'(t - \tau, \tau) d\tau + \int_0^{t+T} F_\tau u'_{t-\tau} d\tau + \int_0^\infty u_{t\tau_1} d\tau_1, \quad (93)$$

$$y' = \int_0^{t+T} N_\theta v'_{t-\theta} d\theta + \int_0^\infty v_{t\theta_1} d\theta_1, \quad (94)$$

$$u'_t = \int_0^{t+T} L_\theta v'_{t-\theta} d\theta, \quad (95)$$

$$v' = \bar{v}' + \bar{v}', \quad (96)$$

where

$$\bar{v}' = \Phi V U \int_0^{t+T} \omega_\tau F'(t - \tau, \tau) d\tau + \Phi V \int_0^{t+T} \omega_\tau F_\tau u'_{t-\tau} d\tau + U G \int_0^{t+T} \Phi_\theta v'_{t-\theta} d\theta,$$

and

$$\bar{v}' = \bar{x} \bar{K} V + \bar{X} \int_0^{t+T} \bar{K}_\theta v'_{t-\theta} d\theta.$$

These five equations contain five unknown functions and are sufficient to determine these functions for all values of $t > -T$. It is often convenient to consider the disturbance as occurring at $t = 0$, in which case the necessary equations are obtained by putting $T = 0$ in the above. It is hoped to discuss the nature of the solutions of these somewhat complicated equations in a future communication.

We shall now consider special examples of the general case corresponding to the five special cases treated on pp. 62 and 63. The results are summarised in Table II. The results in each special case may be verified by working it out *ab initio*, and also by substituting for the variables l_θ , d_θ , etc., the corresponding constant coefficients when the values detailed in Table I are obtained.

(1) $m = 0$, $\bar{\mu} = \mu$, $v = 0$. No immigration; all healthy persons reproduce at a constant rate.

(2) $m = 0$, $\bar{\mu} = \mu$, $v = 0$ and $l_\theta = 0$. As in case (1), but in addition the disease is fatal.

(3) $\bar{\mu} = \mu = v = 0$. Immigration operative but no birth-rate.

(4) $\bar{\mu} = \mu = v = 0$ and $l_\theta = 0$. As in case (3), but in addition the disease is fatal.

(5) $m = \bar{\mu} = \mu = v = 0$ and $d_\theta = 0$. No deaths, no births, no immigration, the total population $n = x + y$. Here we are dealing with a closed population, and with a disease which is never fatal, and which may be recovered from. The total population in the steady state is

$$n = X + Y = V (F(V) + N). \quad (97)$$

The fifth case requires fuller discussion. Let us consider the value of $\chi(V) = V(F(V) + N) - n$ as V tends to zero. Clearly

$$\chi(V)_{V \rightarrow 0} = \text{Limit } \{VF(V)\}_{V \rightarrow 0} - n.$$

We assume that $\bar{\Omega}_\tau \equiv \int_0^\tau \omega_\lambda d\lambda$ has the properties given on p. 70; that is, it is zero between $\tau = 0$ and $\tau = \varepsilon$, then it increases monotonically, and finally for large values of τ greater than η it becomes $\bar{\Omega}_\eta + \omega(\tau - \eta)$.

Thus

$$\begin{aligned} \{\chi(V)\}_{V \rightarrow 0} &= \{VF(V)\}_{V \rightarrow 0} - n, \\ &= \left\{ V \int_0^\varepsilon e^{-\Phi V \bar{\Omega}_\tau} d\tau + V \int_\varepsilon^\eta e^{-\Phi V \bar{\Omega}_\tau} d\tau + V \int_\eta^\infty e^{-\Phi V \bar{\Omega}_\tau} d\tau \right\}_{V \rightarrow 0} - n, \\ &= \left\{ V\varepsilon + V(\eta - \varepsilon) e^{-\Phi V \bar{\Omega}_\eta} + V \int_\eta^\infty e^{-\Phi V [\bar{\Omega}_\eta + \omega(\tau - \eta)]} d\tau \right\}_{V \rightarrow 0} - n, \end{aligned}$$

which, as before, if ε and η are finite

$$\begin{aligned} &= \left\{ V \int_\eta^\infty e^{-\Phi V [\bar{\Omega}_\eta + \omega(\tau - \eta)]} d\tau \right\}_{V \rightarrow 0} - n, \\ &= \left\{ e^{-\Phi V (\bar{\Omega}_\eta - \omega\eta)} V \int_\eta^\infty e^{-\Phi V \omega \tau} d\tau \right\}_{V \rightarrow 0} - n, \\ &= \frac{1}{\Phi \omega} - n. \end{aligned} \tag{98}$$

Further $\{\chi(V)\}_{V \rightarrow \infty} \rightarrow +\infty$, hence if, $\frac{1}{\Phi \omega} - n$ is negative, the equation $\chi(V) = 0$ has one real positive root, and may in general have an odd number of such roots, whilst if $\frac{1}{\Phi \omega} - n$ is positive, there are either no real positive roots, or an even number of such roots. We shall now assume that $d\omega/d\tau$ is never negative. This is the case which is most important in practice, since the condition implies that after an attack of the disease susceptibility remains constant or increases, but never decreases. We shall also assume in the first instance that ω_0 is not equal to zero. We shall now show that under these conditions $\chi(V)$ always increases with V .

$$\begin{aligned} VF(V) &= V \int_0^\infty e^{-\Phi V \int_0^\tau \omega_\lambda d\lambda} d\tau, \\ &= -\frac{1}{\Phi} \int_0^\infty \frac{1}{\omega_\tau} de^{-\Phi V \int_0^\tau \omega_\lambda d\lambda}, \\ &= \frac{1}{\Phi \omega_0} - \frac{1}{\Phi} \int_0^\infty e^{-\Phi V \int_0^\tau \omega_\lambda d\lambda} \frac{1}{\omega_\tau^2} \frac{d\omega}{d\tau} d\tau. \end{aligned} \tag{99}$$

Thus

$$\chi(V) = \frac{1}{\Phi\omega_0} - \frac{1}{\Phi} \int_0^\infty \frac{1}{\omega_\tau^2} \frac{d\omega}{d\tau} e^{-\Phi V \int_0^\tau \omega_\lambda d\lambda} d\tau + NV - n. \quad (100)$$

It is readily seen from the form of this expression that $\chi(V)$ increases with V , provided that $d\omega/d\tau$ is never negative. Let us now suppose that $\omega_\tau = 0$, between $\tau = 0$ and $\tau = \varepsilon$, and let $\omega_{\varepsilon+\delta\varepsilon} \equiv \omega'$, a small but finite quantity. Then

$$\begin{aligned} VF(V) &= V \int_0^\varepsilon e^{-\Phi V \bar{\omega}_\tau} d\tau + V \int_\varepsilon^{\varepsilon+\delta\varepsilon} e^{-\Phi V \bar{\omega}_\tau} d\tau + V \int_{\varepsilon+\delta\varepsilon}^\infty e^{-\Phi V \bar{\omega}_\tau} d\tau, \\ &= V(\varepsilon + \delta\varepsilon) + \frac{1}{\Phi\omega'} - \frac{1}{\Phi} \int_{\varepsilon+\delta\varepsilon}^\infty \frac{1}{\omega_\tau^2} \frac{d\omega}{d\tau} e^{-\Phi V \bar{\omega}_\tau} d\tau. \end{aligned} \quad (101)$$

It follows as above that $\chi(V)$ constantly increases as V increases, even though ω_τ for small values of τ is zero. Thus it appears that if $d\omega/d\tau$ is not negative, the equation $\chi(V) = 0$ has one and only one real root provided that $n > 1/\Phi\omega$, and has no real roots if $n < 1/\Phi\omega$. The value $n_0 = 1/\Phi\omega$ may be called the *threshold density* of population, as no endemic disease can exist when the density is less than $1/\Phi\omega$. Let us suppose that the actual density exceeds the threshold density by a small amount n' , and further let

$$\left\{ \frac{d}{dV} VF(V) \right\}_{V=0} = \lambda.$$

From what has been proved above λ must be positive if $d\omega/d\tau$ is never negative and not constantly zero. If $d\omega/d\tau$ is constantly zero, ω_τ is a constant, and it is then easy to show that λ is equal to zero. Apart from this special case λ is positive. The equation for V then becomes

$$n_0 + n' = \frac{1}{\Phi\omega} + \lambda V + NV + \text{higher powers of } V.$$

As $n_0 = 1/\Phi\omega$, the equation will have a solution V so small that squares and higher powers of V may be neglected. Hence to a first approximation

$$V = \frac{n'}{\lambda + N}, \quad \text{whence} \quad Y = \frac{n'N}{\lambda + N} \quad \text{and} \quad X = n_0 + \frac{n'\lambda}{\lambda + N}. \quad (102)$$

As λ is in general positive, Y is less than n' , and $X - n_0$ is positive. Hence when n' individuals are added to a population already at its threshold density, a certain amount of endemic disease will set in, the number who are ill being less than the actual excess of the population density over the threshold density.

In the equilibrium condition the density of the healthy population will in general exceed the threshold. In the special case where ω is a constant the number who are ill will actually equal the excess, and the number who are well will equal the threshold. The value of λ —apart from simple cases, such as $\omega = \text{constant}$ —is related apparently in a complicated way to the function ω_r , and so far we have been unable to find a simple expression for it.

Discussion.

The mathematical results which we have obtained above are very general in character, and a complete discussion of all the special conclusions which may be drawn would of necessity be unduly long. We shall, therefore, at present confine ourselves to pointing out some of the more important features of the system which is under investigation. Two important limitations of the above theory should in the first place be emphasised. The effect of the age of the individuals is not taken into account at all. Other things being equal, individuals are assumed to have the same susceptibility, the same infectivity and the same chance of dying or recovering whatever their age. In most cases this assumption will not be true. In diseases of comparatively short duration and especially those confined to particular age periods such as childhood, the effect of age may be of relatively little importance, but in diseases which last for periods comparable with the average age during life, the effect of age is likely to be very much more important. The second factor which has been neglected in the above treatment is the effect of death from causes other than the single disease which is assumed to be operative, or the removal of individuals from the population through some independent mechanism. For example, in the case of diseases limited to a particular age group the individual would automatically pass out of the population susceptible to the disease by becoming older. We hope in the future to extend the treatment which has been elaborated in the present paper so as to take into account the effect of a general death rate, and if possible the effect of age.

The results obtained above refer partly to the conditions for the existence of a steady state, and partly to the nature of the equilibrium so obtained. It is of considerable interest that in the general problem a particular system admits at the most of only one steady state provided that the susceptibility never decreases as time goes on. This follows from the fact that equation (63) has only one real positive solution. It may have no real positive solution at all, in which case no steady state can exist. Equation (68) gives the condition under

which this will occur. This limitation of the number of steady states to one at most, is not immediately obvious. One did not know for certain that a number of steady states might not exist, each being characterised by a different value of V .

In the problem dealt with in our previous paper we showed that a threshold population density existed which was such that no epidemic could occur if the population density was less than this threshold. In the present problem a threshold of an analogous type turns up only in case 5, that is, when the population is closed so that no individuals can leave or enter it. In the other cases the population level automatically adjusts itself in conformity with the values of the arbitrary functions. Naturally no such adjustment is possible with the closed population considered in case 5, and it has been shown that in this instance no disease can exist unless the total density of population exceeds a certain level. In the special case of constant coefficients this threshold density happens to be identical with the threshold density found in dealing with constant coefficients in the previous paper, being equal to $1/k$. Further, we have also shown that in the case of constant coefficients the number of healthy individuals remains at the threshold level when the number of the population exceeds this threshold, and that the excess are found in the diseased group. In the general case, however, this does not necessarily hold, the number who are ill may not, and in general will not, be equal to, but form only a fraction of the excess, so that the number of healthy individuals in the population will usually become greater as the total population increases beyond the threshold. The disease rate of the population will, to a first approximation, be proportional to the excess of the population over the threshold, when this excess is small.

Some interesting results follow from the present work in regard to the effect of the rate of influx of fresh individuals on the mortality rate, and the endemic level of a population. The mortality rate is equal to $W/n = DT$, the disease rate is denoted by $V/n = T$, whilst the endemic level or fraction of the population actually infected is $Y/n = NT$. We have shown that an increase in the immigration rate, or in the birth-rate either of the virgin, the recovered or the diseased, will result in an increase of T , and therefore in increases of the mortality rate, the disease rate and the endemic level. A fall in immigration rate or birth-rate will have the opposite effect. Although in obtaining this result the various assumptions mentioned at the beginning of the discussion have been made, yet it is to be expected that in general a fall in the birth-rate would be accompanied by a fall in the prevalence of all contagious diseases. We hope to deal with this problem more fully at a later date, when considering

a system in which allowance is made for deaths or removals from causes other than the single disease under discussion.

The equilibrium states which we have been discussing so far may be either stable or unstable. In the former case, when the population is almost, but not quite, in equilibrium, then as time goes on it will gradually approach the equilibrium condition. When the equilibrium is unstable one or two things may happen. The population may gradually move farther and farther away from the equilibrium condition, until the disease completely disappears, or the population is wiped out. On the other hand, the population may oscillate about its equilibrium state, but these oscillations may gradually increase in amplitude until they become very great. (As a border line case we have the special possibility of continuous oscillations of constant amplitude, but this will appear only under very special conditions.) We may define the first type of instability as aperiodic, and the second as periodic. The nature of the equilibrium in the case of constant coefficients will be determined by the nature of the roots of the equation (14) in α . If one of these roots corresponds to unstable equilibrium, then, although the others may correspond to stability, yet the state will be unstable. In the same way in an unstable system the instability might in general be mixed, that is, the equation in α might have roots corresponding to both types of instability. It is readily seen that the condition for stability is that the real roots of the equation should be negative and that the complex roots should have negative real parts. We have shown that in general this condition is always satisfied in the case of constant coefficients. The only exception occurs in the limiting case (3) in which the real part is zero, and pure harmonic vibrations result. The exceptional value $\alpha = 0$ obtained in case (5) accommodates a disturbance within a closed population resulting in a change in the total number.

The theory as elaborated in this paper only applies to oscillations of small amplitude about the equilibrium condition. When these amplitudes increase the theory will cease to hold. It does not give a complete account of what happens in a system possessing periodic or aperiodic instability. It only indicates the nature of the stability. As, however, there is at most only one equilibrium condition the system cannot change to any steady condition other than one in which there is no disease at all or one in which the whole population has been wiped out. Apart from these two possibilities, it seems that it is not possible from the above to indicate how a system exhibiting periodic instability may behave after the oscillations have reached a considerable amplitude.

Summary.

(1) A mathematical investigation has been made of the prevalence of a disease in a population from which certain individuals are being removed as the result of the disease, whilst fresh individuals are being introduced as the result of birth or immigration. Allowance is made for the effects of the immunity produced as the result of an attack of the disease, but the effect of deaths from other causes is not taken into account, and the action of the disease is supposed to be independent of the age of the individual.

(2) As a special case of the above, results have been obtained for a closed population in which no deaths occur and to which no fresh individuals are added, but in which the individuals after being infected acquire immunity, and then may be again infected. A threshold density of population exists analogous to that described in the previous paper, which is such that no disease can exist in a population, the density of which is below the threshold.

(3) In other special cases investigated when either immigration or birth is operative in the supply of fresh individuals, as well as in the general case, only one steady state of disease is possible. To reach this state the population must be of a certain density which will be determined by the functions characterising the infectivity, morbidity, etc., of the disease.

(4) Increase of the immigration rate or of the birth-rate results in an increase in the rate of infection of the healthy individuals and also in the percentage rate of infection, the percentage of sick, and in the percentage mortality from the disease. This result is, of course, a necessary consequence of our assumption that the disease is the only cause of death.

(5) More particular results have been obtained by substituting constants in the place of the undetermined functions assumed in the general theory. Further, under these conditions the nature of the steady states has been more fully investigated and it has been shown that in all cases, except one, the steady states are stable ones. In the exception, a disturbance would result in purely periodic oscillations about the steady state.

*The Absorption Spectrum of Nitrous-Oxide and the Heat of
Dissociation of Nitrogen.*

By ARUN K. DUTTA, M.Sc., Department of Physics, Allahabad University,
India.

(Communicated by M. N. Saha, F.R.S.—Received May 9, 1932.)

In some previous papers it has been shown by the author and others* that saturated compounds of most substances in the vapour state show continuous absorption. A typical example is SO_2 -vapour, which was recently studied by the author† and which enabled him to make an accurate estimation of the heat of dissociation of oxygen. In the present work, the absorption spectrum of N_2O was investigated with a view to determining the heat of dissociation of nitrogen.

Leifson‡ was the first to investigate the absorption spectrum of N_2O gas and found that the gas shows no selective absorption in the Schumann region. He states that the absorption is in the form of two continuous bands, the first extending from λ 2000 to λ 1680 and the second from λ 1550 beyond the range of observation. Recently Wulf and Melvin§ showed that when N_2O is illuminated with light of wave-length λ 2300, it is decomposed photochemically into NO and N; they also noticed that N_2O possesses no band absorption.

Experimental Work.

As the previous data appeared to be fragmentary, it was decided to investigate the spectrum in greater detail. The gas was prepared by heating pure NH_4NO_3 and purified by passing through ferrous sulphate solution. It was collected in a glass tube 100 cm. long, whose ends were closed by quartz windows. The source of continuous light was a hydrogen discharge tube provided with thick aluminium electrodes and run from a 2-kilowatt transformer. It gave a perfectly continuous spectrum from extreme red right up to the limit of quartz. The spectrum was photographed in a small quartz spectrograph.

* Dutt, Saha and Deb., 'Bull Acad. Sci. Allahabad,' vol. 1.

† 'Proc. Roy. Soc.,' A, vol. 137, p. 366 (1932).

‡ 'Astr. J.,' vol. 63, p. 73 (1926).

§ 'Phys. Rev.,' vol. 39, p. 180 (1932).

It was observed that N_2O gas gave no band absorption. Light appeared to be absorbed from about λ 2400, but as the position of the cut seemed to shift with pressure, it was apparent that the eye-estimation of the cut was likely to lead to grave errors. A detailed microphotometric study of the intensity of the plates obtained under different pressures was therefore undertaken.

The microphotometer used was of the Zeiss-pattern, belonging to the physics department of the Patna University, and I wish to express my thanks to Professor A. T. Mukherjee, Director of the Patna Laboratory, for his kind permission to use this apparatus. The intensities of the continuous spectrum as well as that of the spectrum transmitted through N_2O under different conditions were measured throughout the entire range of wave-lengths and the transmission ratios for each wave-length was obtained by comparing the intensity of the spectrum transmitted through the gas with that of the spectrum transmitted through the empty tube. It may be added that the time of exposure as well as the conditions of running were maintained as constant as possible. These transmission ratios are shown in Table I.

Table I.

Circumstances under which the spectrum was taken.	Percentages of transmitted light at wave-lengths.					
	λ 2618.	λ 2442.	λ 2370.	λ 2294.	λ 2215.	λ 2149.
Spectrum transmitted through 100 cm. of N_2O at atmospheric pressure	93.75	86.25	81.12	60.0	35.0	8.75
Spectrum transmitted through 100 cm. of N_2O at half atmospheric pressure	96.25	90.62	84.38	71.25	49.38	25.75
Spectrum transmitted through 100 cm. of N_2O at quarter atmospheric pressure	97.75	92.50	86.87	75.0	56.25	37.50

From this table a curve was plotted with the percentage of absorption (absorption ratio = $1 - \text{transmission ratio}$) as ordinate and wave-lengths as abscissæ. The curves for the three different conditions are shown in fig. 1. It is seen that all the three curves cut the abscissæ at one and the same point. This point must, therefore, be taken as marking the beginning of absorption. From the graph, the position of the beginning of absorption was found to be at λ 2750.

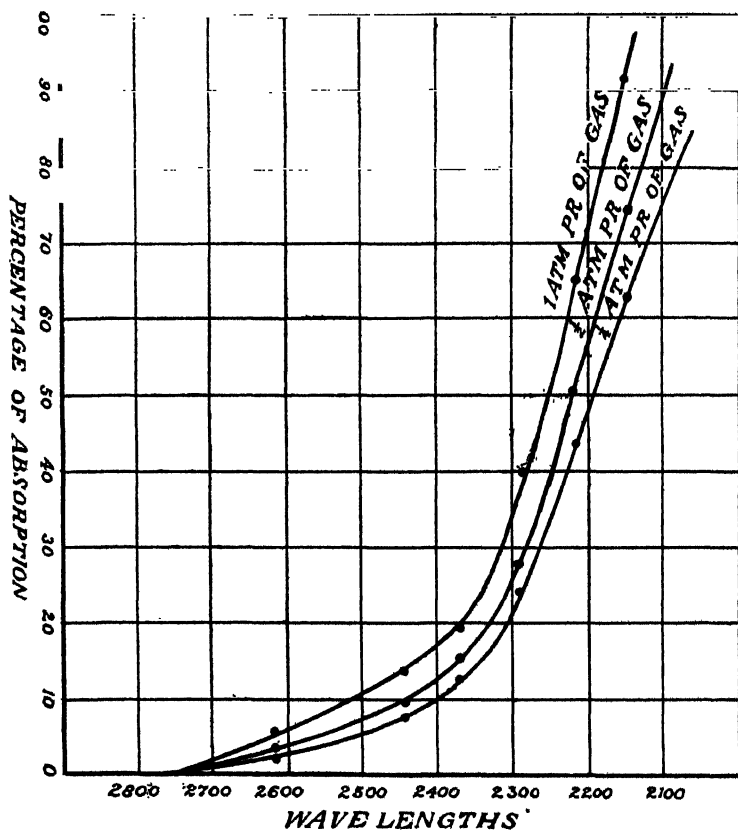
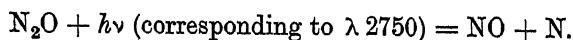


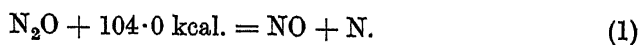
FIG. 1.

Discussion of Results.

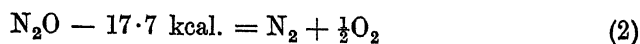
From the above spectrophotometric work it has been established that N_2O gas begins to absorb light from λ 2750. As the absorption is continuous we can suppose that the photochemical reaction takes place as follows :—



Now λ 2750 corresponds to 104.0 kcal. We have, therefore,



Again, from thermochemical data*



* Landolt and Bornstein's Tables, p. 1489 (1923).

and

$$\text{NO} - 21.5 \text{ kcal.} = \frac{1}{2}\text{N}_2 + \frac{1}{2}\text{O}_2. \quad (3)$$

From (1), (2) and (3), we have

$$\frac{1}{2}\text{N}_2 + 100.2 \text{ kcal.} = \text{N}$$

or

$$\text{N}_2 = 2\text{N} - 200.4 \text{ kcal.},$$

i.e.,

$$D_{\text{N}_2} = 200.4 \text{ kcal.} = 8.7 \text{ volts.}$$

This value of D_{N_2} seems to be the most reliable so far obtained ; it is discussed later.

In the above discussion it has been assumed that N_2O decomposes into normal NO and normal N. In the case of SO_3 , it was found that after a first absorption at λ 3300, which corresponds to dissociation of SO_3 into SO_2 and normal O, there is a patch of transmitted light followed by a second absorption at λ 2300 ; this was interpreted as marking the decomposition of SO_3 into SO_2 and an excited oxygen atom in the $^1\text{D}_2$ state. It is very probable that N_2O should show a similar behaviour. The absorption at λ 2750 has been interpreted to indicate dissociation of N_2O into NO and normal N-atoms. Leifson's (*loc. cit.*) investigations show that the first absorption is followed by retransmission, which is again followed by a second absorption which becomes prominent at λ 1550. It is very probable that the second absorption indicates photochemical decomposition of N_2O into NO and excited N-atoms. Now, the fundamental state of N-atom is $^4\text{S}_2$ and the next metastable state is ^2D . According to Compton and Boyce* $^4\text{S}_2 - ^2\text{D}$ is for nitrogen very nearly 2.37 volts. From these data we can calculate the expected beginning of second absorption. We get,

$$\text{N}_2\text{O} + 6.91 \text{ volts} = \text{NO} + \text{N}(^2\text{D}) \quad (4)$$

$6.91 \text{ volts} = 4.54 + 2.37 = \text{volt corresponding to } \lambda \text{ 2750} + \text{volt corresponding to } ^4\text{S}_2 - ^2\text{D} \text{ of N.}$ This corresponds to $\lambda = 1840 \text{ \AA}$. This apparently does not agree with Leifson's value, but as I have already pointed out here and also in a separate paper,† the eye-estimation of the beginning of absorption is likely to give rise to grave errors. For example, according to eye-estimation, absorption of N_2O appears to begin at λ 2400, whereas the microphotometrical studies of the absorption spectrum shows that it begins at λ 2750. It is therefore very probable that if the absorption spectrum

* 'Phys. Rev.', vol. 33, p. 145 (1929).

† Dutt, 'Z. Physik' (in course of publication).

of N_2O be studied again in a fluorite or vacuum spectrograph, the second cut will be found nearer the expected region.

Discussion on the Heat of Dissociation of N_2 .

A resumé of the investigation on the heat of dissociation of N_2 is given in Table II ; of these, only a few merit discussion.

Table II.—Heat of Dissociation of N_2 .

	Method.	Value in volts.
Eucken*	Thermal.....	19.1 to 16.5
Langmuir†	".....	>10
Sponer‡	Active nitrogen.....	11.4
Cario and Kaplan§	".....	9.6
Gaviola	Sensitised fluorescence.....	9.3 or 9.8
Sponer¶	Band spectrum data: extrapolation of N_2^+ bands....	11.75
Herzberg**	Discussion of N_2^+ bands.....	9.1
Birge††	".....	9.5
Mulliken‡‡	General criticism.....	9.5
Birge§§	".....	9.0
Grimm	Electron bombardment.....	15.9
Tate and Lozier¶¶	".....	8.4
Turner and Samson***	Excitation potential of the negative bands of N_2^+	8.4

* 'Leibig. Ann.,' vol. 440, p. 111 (1924).

† 'J. Amer. Chem. Soc.,' vol. 37, p. 417 (1915).

‡ 'Z. Physik,' vol. 34, p. 622 (1925).

§ 'Nature,' vol. 121, p. 906 (1921).

|| 'Nature,' vol. 122, p. 313 (1928).

¶ 'Z. Physik,' vol. 41, p. 611 (1927).

*** 'Phys. Rev.,' vol. 34, p. 747 (1929).

** 'Nature,' vol. 122, p. 505 (1928).

†† 'Nature,' vol. 122, p. 842 (1928).

‡‡ 'Nature,' vol. 122, p. 842 (1928).

§§ 'Phys. Rev.,' vol. 34, p. 1062 (1929).

||| 'Z. Electrochemie,' vol. 31, p. 474 (1925).

¶¶ 'Phys. Rev.,' vol. 39, p. 254 (1932).

In the work of Tate and Lozier, N_2 gas was bombarded by electrons of varying velocity and the characteristics of the resulting products were studied by a mass spectrograph. It was found that the N^+ ions were just produced when the voltage of the bombarding electrons reached the value 22.9 ± 0.5 volts. The only interpretation of this result is that the N_2 -molecule is split up into normal N, normal N^+ , and an electron, when it is hit by an electron with the energy 22.9 ± 0.5 volts. We have, therefore,

$$N_2 = N(^4S) + N^+(^3P) + e - (22.9 \pm 0.5) \text{ volts.} \quad (5)$$

But,

$$N(^4S) + 14.5 \text{ volts} = N^+(^3P) + e$$

as determined by Hopfield.* Hence

$$N_2 = N(^4S) + N(^4S) - (8.4 \pm 0.5) \text{ volts.}$$

* 'Phys. Rev.,' vol. 127, p. 801 (1926).

This method is very straightforward, but the authors point out that owing to certain unavoidable circumstances of the experiment the accuracy cannot be much increased, and the result is uncertain to the extent of ± 0.5 volts. This value is in good agreement with that obtained by the writer.

The other set of values is grouped round the fundamental work of Herzberg on the band structure of N_2^+ . From a study of the band spectrum of N_2^+ Herzberg* came to the conclusions that the N_2^+ normal molecule is formed of a normal N^+ (3P) atom and an excited N (2D) atom,† while in the excited state it consists of normal N^+ (3P) and normal N (4S) atoms as shown in fig. 2.

These facts can be represented as follows :—

$$N_2^+ (\text{normal}) = N (^2D) + N^+ (^3P) - 9.1 \text{ volts} \quad (6)$$

$$N_2^+ (\text{excited}) = N (^4S) + N^+ (^3P) - 3.7 \text{ volts.} \quad (7)$$

The values of 9.1 and 3.7 were obtained by Birge and Spooner's method from the vibrational levels of N_2^+ in the normal and excited states. Again,

$$N_2^+ (\text{excited}) = N_2^+ (\text{normal}) + 3.2 \text{ volts,}$$

3.2 volts corresponding to the transitions of the (0—0) band of N_2^+ . Hence

$$N_2^+ (\text{normal}) = N (^4S) + N^+ (^3P) - 6.9 \text{ volts.}$$

Herzberg further used the relation

$$D_{m'} + I_m = D_m + I_a \quad (8)$$

where

I_a = heat of ionisation of N-atom from

$$N (^4S) \text{ to } N^+ (^3P) = 14.5 \text{ volts.}$$

D_m = heat of dissociation of N_2 molecule.

I_m = heat of ionisation of N_2 molecule.

$D_{m'}$ = heat of dissociation of N_2^+ molecule into normal states of the atoms, *i.e.*, $N (^4S)$ and $N^+ (^3P)$.

* 'Ann. Physik,' vol. 86, p. 189 (1928).

† Regarding the interpretation of the bands of N_2^+ and other similar compounds, CN, SiN, CO⁺, etc., the present position seems to be less clear than was first supposed by Herzberg. While the excited states in all these compounds are assumed to be made up of normal atoms, no uniformity is obtained as regards the state of atoms composing the normal molecule when detailed numerical comparisons are made. For discussions, Herzberg and Heitler, 'Z. Physik,' vol. 53, p. 53 (1929). But for evaluating the heat of dissociation of N_2 these controversial points are not involved. We have to use only equation (7).

From the older experiments on the ionisation potential of molecules Herzberg took I_m to be 16.7 volts, but this value is uncertain. D_m was taken to be 6.9 volts. Substituting these values in (8) D_{N_2} was found to be 9.1 volts. The accuracy of this value depends upon the values of D_m and I_m .

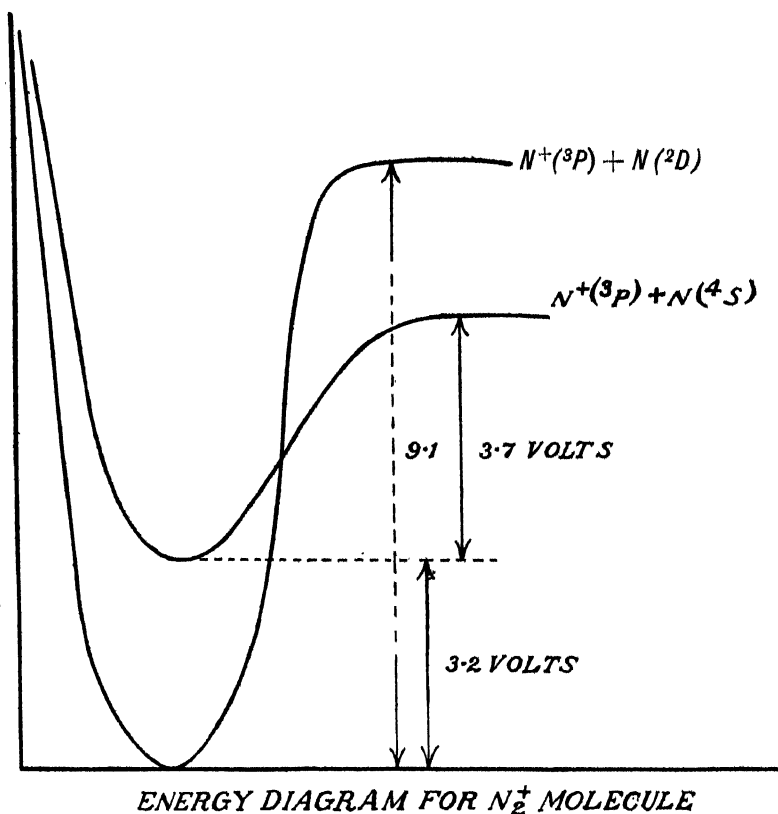


FIG. 2.

(The abscissa represent distance between the constituent nuclei, the ordinate, the potential energy in volts. The other points are discussed below.)

In the experiments of Turner and Samson the minimum electron voltage for getting the (0—1) band of N_2^+ was found to be 19 volts. We have therefore,

$$\begin{aligned} N_2 &= N_2^+ (\text{excited}) + e - 19 \text{ volts} \\ &= N_2^+ (\text{normal}) + e - 15.8 \text{ volts} \end{aligned}$$

(excitation energy of the (0—0) band being 3.2 volts)

$$= N(^4S) + N^+(^3P) + e - 22.7 \text{ volts}$$

from Herzberg's value of 6.9 volts for D_m .

But $N^+ (^3P) = N (^4S) - e + 14.5$ as obtained by Hopfield (*loc. cit.*). Therefore $N_2 = 2N (^4S) - 8.2$ volts. This value is subject to two uncertainties: firstly, in the measurement of the excitation potential for raising N_2 to N_2^+ . This is given as 19 volts, but the actual value obtained was 22 volts. The difference is the correction for contact potential which has been obtained by comparison of the excitation potential of the line 6402.2 of Ne. Secondly, in the assumption that the normal state is composed of all the molecules in the zero vibrational state. Birge* also drew attention to this point.

Gaviola observed that the NH band appears as a result of photosensitised fluorescence when about 4 mm. of nitrogen and very little hydrogen are admitted to the quartz tube containing the mercury vapour, which is being excited by the light of a water cooled, magnetically deflected mercury arc. The NH band disappears as soon as the line 2537 of the arc is absorbed or self reversed. Having determined that the intensity of the bands is proportional to the square of the intensity of the exciting light, he concludes that atomic nitrogen is formed by three body collisions of N_2 molecules with two excited mercury atoms. According to whether the mercury atoms are excited to the 3P_1 state with 4.9 volts of energy, or to metastable 3P_0 state with 4.68 volts, the value of D_{N_2} obtains an upper limit of 9.8 volts or 9.3 volts. The values are too high and it is probable that in this case dissociation takes place with considerable kinetic energy.

In conclusion I wish to thank Professor M. N. Saha, F.R.S., for his kind interest and valuable suggestions during the course of this work.

Summary.

(1) The absorption spectrum of N_2O has been observed in the quartz region. Continuous absorption was obtained and microphotometer measurements were taken to locate accurately the beginning of absorption.

(2) The beginning of absorption takes place at λ 2750, corresponding to 104.0 kcal. Assuming the process of decomposition to take place according to the equation $N_2O + E = NO + N$, the heat of dissociation of N_2 has been obtained as 8.7 volts, with the help of other thermochemical data.

(3) A second reabsorption is expected near λ 1840 Å, and it is thought probable that the apparent reabsorption obtained by Leifson at λ 1550 would shift further to the red side if microphotometer measurements were taken.

(4) Other methods of evaluating D_{N_2} have been reviewed and criticised.

* 'Phys. Rev.', vol. 34, p. 1062 (1929).

The Mechanism of the Initiation and Propagation of Detonation in Solid Explosives.

By WILFRID TAYLOR, Ph.D., and ALFRED WEALE, M.Sc., Nobel House,
Imperial Chemical Industries, Ltd., Stevenston, Ayrshire.

(Communicated by F. G. Donnan, F.R.S.—Received May 20, 1932.)

Although half a century has elapsed since the publication of the classical treatise of Berthelot upon explosives, the detailed mechanism of the initiation and propagation of detonation in liquid and solid explosives is still obscure. Detonation is a phenomenon exhibiting a number of specific characteristics which differentiate it quite definitely from the explosive combustions of such substances as gunpowder and cordite. It is well known that the latter are governed by laws relating the rate of reaction to the surface area, the temperature and pressure of the surrounding gases, etc., and that heat is the chief medium of initiation and propagation, whereas in the case of detonation, the reaction wave-front travels directly through the explosive medium in the same sense as does a sound wave, and the velocity of propagation is a very definite characteristic of the phenomenon. This stability of the detonation velocity is well demonstrated for solid explosives by the photographs in a recent paper by E. Jones*; the speeds are usually much greater than any exhibited by explosive combustions, and range from 1500 to 10,000 metres per second. Finally, the initiation and propagation of detonation appear to be associated much more intimately with mechanical shock than with flame. The weight of evidence strongly indicates that the difference between detonation and explosive combustion is fundamental and not merely of degree, and the term "high explosive" is reserved for substances capable of the former property.

The theoretical treatment of detonation as a shock wave traversing the medium and maintained by the accompanying chemical reactions has been developed by several investigators.† These writers have built up a quantitative theory from thermodynamical reasoning and have been able to calculate velocities of propagation, which in some cases are correct, but in practice it has been found that the thermodynamical conditions, while necessary, are not sufficient. Thus, a great number of compositions possessing all the

* 'Proc. Roy. Soc.,' A, vol. 120, p. 603 (1928).

† See Jouguet, "Mecanique des Explosifs."

thermodynamical qualifications of a high explosive cannot be made to detonate ; others permit detonation to be initiated successfully but without propagation, and the reaction degenerates into a mere deflagration, or even dies out completely. It is indeed very difficult to judge whether a particular composition is a true detonating explosive without the opportunity to test the sample in reasonable quantity. The violent decompositions of small samples or single crystals furnishes no *a priori* evidence of detonation, and innumerable examples may be quoted of such material in bulk being unable to propagate the local and violent initial activity.

The chief reason for the insufficiency of the classical mechanical theories to forecast the behaviour of a potential high explosive is that they neglect, or do not deal adequately with, the mechanism whereby the pressure or other factors bring about the decomposition of the explosive immediately in advance of the wave-front. Many writers have followed Berthelot who believed that detonation was an internal combustion brought about by the rapid adiabatic compression of each layer of the explosive in turn, the heat developed being sufficient to raise the material to the ignition temperature. The adiabatic compression hypothesis has in fact met with much success in explaining detonation waves in gases, but the validity of its extension to liquid and solid explosives has been doubted by reliable thermochemical authority.*

Reference may also be made to a hypothesis advanced by Abel, who supposed the disintegrative impulse to be conveyed from molecule to molecule by "sympathetic vibrations," the nature of which could not at that time be specified. This theory has been discarded, and certain experimental evidence cited by Abel has since been shown to be inadequate ; but in a slightly more modern garb, the same idea has recently been advanced in another form, the "sympathetic vibrations" being identified with characteristic infra-red frequencies which resonate the unstable molecules selectively. It must be pointed out, however, that continuity of the explosive component is not necessary in a detonating composition, and the detonation wave may even successfully traverse an inert layer intersecting the cartridge. The distinction must again be drawn between the explosive decomposition of a crystal in

* Nernst, "Theoretical Chemistry," p. 789 of 1923 edition, says : "The theory of solid and liquid explosives will probably progress along the same lines as that of gaseous explosives ; in particular, the pressure necessary to produce the explosion will be of fundamental importance for the theory of the development of the explosion wave. It must, however, be observed that, contrary to opinions sometimes expressed, a high temperature such as would initiate the reaction will certainly not be obtained by compression as is the case with gases."

which there is molecular contiguity, and the propagation of the detonation wave through a possibly heterogeneous medium.

After reviewing the literature of the subject, very briefly touched upon in the above outline of the problem, it was felt that the chief need in this field was the prosecution of quantitative experimental work and the accumulation of reliable data before proceeding to elaborate further hypotheses. A general investigation was therefore planned, and a description of the preliminary results forms the substance of the present paper.

The treatment of the problem adopted in this research was as follows. As the detonation wave-front advances across any thin layer of the medium, three main features are to be distinguished as contributing to the propagation, namely (*a*) the pressures and stresses set up either by the original initiating impulse or by the decomposition products of the preceding layers; (*b*) the physical and chemical characteristics of the layer itself, conveniently designated by the term "sensitiveness" and which determine its readiness to decompose; and (*c*) the mechanical reaction of the undecomposed material immediately in advance of the wave-front. Previous investigations have been concerned principally with (*a*) and information upon (*b*) has been greatly lacking, although it would appear that the real key to the problem is to be found in this direction.

The preliminary experimental work was concerned with an investigation of the conditions whereby a thin layer of explosive can be initiated to explosive decomposition by percussive forces of the same order of magnitude as those present in the shock waves associated with detonations. If the conditions governing the initiation of the first layer be known, then the natural assumption is to be made that the next layer is initiated in turn by the transmitted portion of the original blow plus the contribution due to the pressure developed by the explosion products of the first layer.

The pressures involved in detonation are extremely high, of the order of 100 tons per square inch and upwards. The only possibility of working with pressures of the right order lay therefore in selecting the most sensitive substances, and employing the great forces called into play during the impacts of hard metals. An existing type of sensitiveness test, namely the "fall hammer" or "drop-weight" test, was therefore made the starting-point of the investigation. Unfortunately, the results of this test are notoriously lacking in reliability and great care was required in the first place to obtain the conditions of successful experiment.

Apparatus and Manipulation.

In the ideal fall-hammer test, a pellet of explosive composition would be placed upon a hard flat surface and be struck by an inelastic falling weight. The variable factors are then the mass of the striker W , the geometry of the striking surface and the height of fall H .

In practice, however, it is difficult to ensure that a relatively large mass will fall squarely and in exactly the same way every time, and hence it has been found desirable to interpose a "striking pin" or piston between the composition and the falling weight. This piston receives the blow and transmits it to the pellet, and it is possible to place and guide the piston so as to obtain much better reproducibility of the blow.

Two types of apparatus were used, similar in principle, one of which was designed for use with very sensitive explosives for which the falling weights ranged from 1 to 64 ounces, and the other of a more sturdy construction with weights ranging from $\frac{1}{2}$ to 20 kilograms falling from heights up to 2 metres. The details of these two instruments will be seen by reference to fig. 1*a*.

The pedestal and framework of the smaller apparatus were constructed of cast iron and weighed approximately 1 cwt. In essence this apparatus consisted of an iron table and a vertical rod, the latter carrying a sliding arm. The striker weights were held at any desired height above the table by means of an electromagnet attached to the sliding arm. Various types of weights were tried, and it was found that the best for the purpose were Hoffmann steel balls, which were readily replaceable and could be obtained in a wide range of sizes. They had the further advantage of being manufactured to specification of high precision, were accurately spherical and of very hard steel.*

The striking pins were of hardened steel, and of various shapes and sizes. They were constrained by the collar arrangement shown in the drawing so as to have freedom of motion in a vertical direction only. In certain cases it was found convenient to replace the cylindrical pistons by small steel balls, 3 or 4 mm. in diameter, as shown in fig. 1, *d*.

Regarding the larger instrument, many forms have been described in the literature.† In the particular form of apparatus used in this work, a steel hammer was constrained by vertical guides to fall upon a massive steel anvil. The sample of explosive and its system of pistons and constraints as shown in

* Steel balls have been used previously by Dupré (7th International Congress of Applied Chemistry, Section IIIb, p. 132).

† Kast, International Congress of Applied Chemistry, 1909, also Lenze, *ibid.*, 1906.

fig. 1, *c*, were placed on the anvil to receive the blow. The steel cylinders were Hoffmann rollers and could be replaced as frequently as desired. Only with the most powerful explosives, however, were frequent replacements

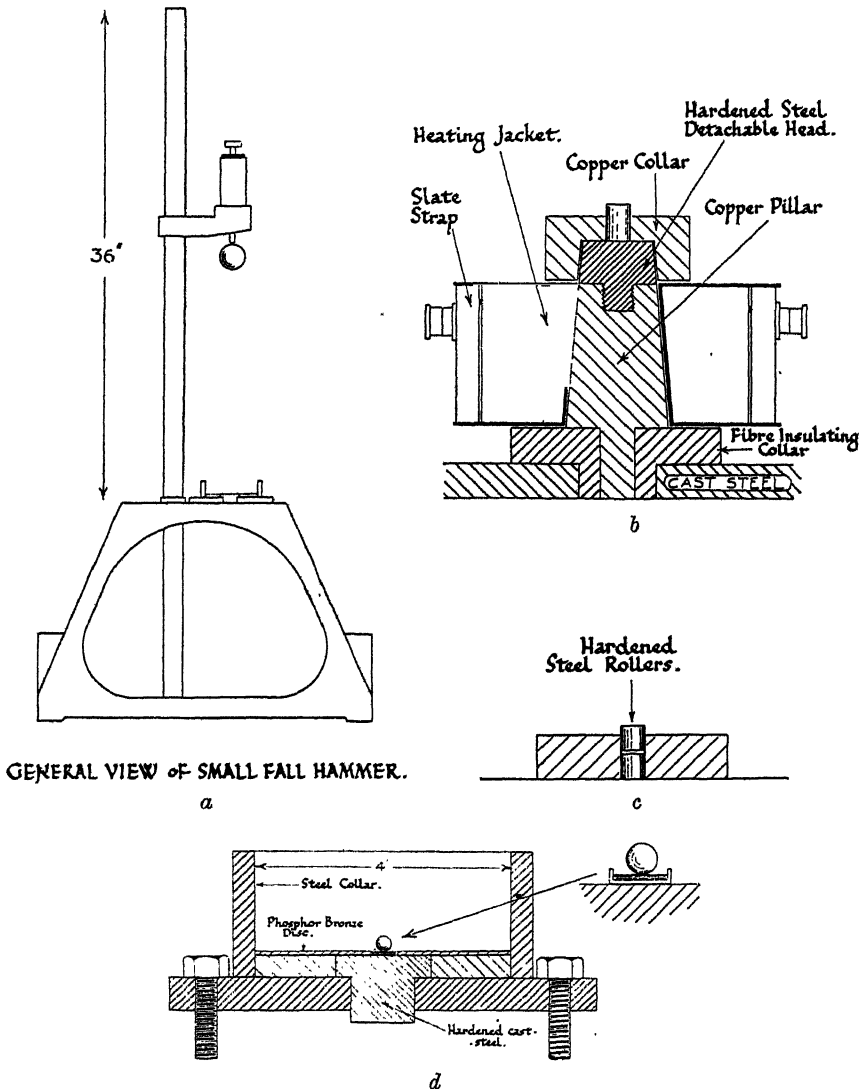


FIG. 1, *a*—*d*.

necessary. The explosive was placed evenly on the upper surface of the lower piston, and the second piston pressed down on it, the two being held in position and centred by the steel collar as shown.

The explosives used were not, in general, pure substances. Detonating

explosives are rarely pure compounds, but contain additional substances added for stability, energy content, flame-quenching properties, etc., and their detonating properties are characteristic of the whole incorporation, so that a general investigation must not be restricted to one component system. The explosives used on the smaller apparatus were chiefly mercury fulminate, lead azide, lead dinitroresorcinate and mixtures of these substances with oxidising materials such as potassium chlorate and barium nitrate, and fuels such as antimony sulphide. On the larger instrument they included tetryl, TNT, picric acid, nitroglycerine gelatines, dynamites, and typical high explosive mixtures. For work on the smaller apparatus the compositions were loaded into copper or brass shells, the charges being about 0.03 gram and pressed into firm pellets under loads varying from 150 to 450 lbs., corresponding to pressures from 6000 to 18,000 lbs. per square inch. The thicknesses of the pellets were of the order of 0.02 inch and the diameters were 0.25 inch. With all explosives used on the two instruments, the criteria of ignition under shock of the falling weight were sufficiently marked to eliminate ambiguity. A very sharp report was obtained, and in many cases the violence of the explosion was sufficient to throw the piston and weight to a distance. It is necessary to stress this point because with other forms of the apparatus used in the preliminary work, very great ambiguity arose owing to incomplete ignition of the less sensitive explosives such as trinitrotoluol-ammonium nitrate mixtures. This difficulty has been commented upon by other workers, but the use of the constraining system described above practically eliminated the trouble. Only complete ignitions were recorded as "fires." In both the large and small instruments the surfaces of the striking pins became pitted and eroded with the violence of the ignitions, but the standard rollers and balls were renewed immediately the surface was attacked.

For a series of experiments with any particular weight and height of fall, both ignitions and non-ignitions are usually obtained. It is usual therefore to define the characteristic height for a given sample as that at which a certain fraction of ignitions occur out of a given number of trials, each pellet being, of course, submitted once only to the blow. It soon became evident, however, that no high accuracy could be obtained by this method except in a few very favourable cases.

It is not surprising that trouble is experienced in obtaining regular results when it is remembered that the experiments deal with very small quantities of mixed materials. It was found necessary therefore to investigate the distribution of ignitions for a standard explosive, using a standard weight and

varying the height of fall only. A typical set of results is given in the following table, and similar distributions were obtained with the larger apparatus and for all the conditions studied.

Table I.—Distribution of Ignitions according to Height of Fall of Striking Weight.

Striking pins : 4 mm. steel balls.

Containing shell : Brass 0.22 inch diameter.

Composition : Mercury fulminate mixture containing ground glass.

Charge : 0.034 gram. Striking weight : 2 ounces.

Thickness of composition : 0.02 inch.

Height of fall, inches (H).	Total number of trials.	Number of ignitions.	Percentage of ignitions (y).
6.0	12	12	per cent. 100
5.5	12	12	100
4.5	12	12	100
4.0	20	20	100
3.5	30	25	83.3
3.0	30	18	60.0
2.7	30	11	36.7
2.5	30	13	43.3
2.0	30	8	26.0
1.5	20	0	0

The results obviously suggest a statistical distribution and suggest the following behaviour. For any given weight and type of explosive, assuming all the samples to be exactly alike, and all struck in an identical manner, there would be a critical height of fall, lesser heights producing no ignitions and greater heights resulting in 100 per cent. ignitions. Actually, however, the samples differ from the ideal one in shape, size, homogeneity, etc., so that each sample will have a critical height peculiar to itself and differing from the ideal one by a positive or negative divergence. The distribution of these divergences will be governed by some probability law.

From a practical point of view it seemed likely that the distribution would be heavily biased on the side of positive divergences, *i.e.*, that it would be much easier to lose than to gain efficiency in the blows, but a preliminary inspection of the results did not support this, and it appeared that a purely random distribution of divergences would fit the facts better. If the divergences are purely random they will be governed by the Maxwell function

$$dN = \frac{h}{\sqrt{\pi}} e^{-h^2 z^2} \cdot dz,$$

where z is the divergence and the other terms have their usual meaning in probability theory. The critical height for any samples taken at random is $H = H_c \pm z$, where H_c is the ideal critical height.

The number of samples having their critical heights less than some given height H will therefore be

$$N = \frac{1}{2}N_0 (1 \pm \text{erf } z),$$

where the sign in the bracket follows the sign of z itself and

$$\text{erf } z = \frac{2h}{\sqrt{\pi}} \int_0^z e^{-h^2 z^2} \cdot dz.$$

Since all those samples having lower critical heights will fire for a height H , the percentage of ignitions will therefore be

$$y = \frac{50N}{N_0} (1 \pm \text{erf } z).$$

This is therefore the theoretical equation to the curve which should be obtained from the results of Table I, assuming the random law of distribution to be correct. The scale of the curve is fixed in the y direction by the value of the modulus h . In practice the experimental points were plotted first and a smooth curve drawn through them. The gradient at the point $y = 50$ per cent. is given by

$$\tan A = 100h/\sqrt{\pi},$$

and was obtained graphically. The theoretical curve was then computed and added to the diagram.

It soon became evident that these curves derived from the Maxwell function were also the best possible fits to the experimental points, any variations being irregular and due presumably to the fact that sufficient numbers of trials had not been made. In order to test this point, a very extensive series of trials was carried out, taking eight heights of fall, and testing 100 samples at each height. These results are shown in fig. 2 together with the theoretical curve plotted from the above distribution function. It is evident that there is no suspicion of any "skew-distribution."

The standard deviation D of the distribution may be obtained by $h = 1/\sqrt{2D}$. It was very necessary to keep D as low as possible; experience with older types of apparatus showed that a high scatter-value was the chief reason for the uncertainty and lack of reproducibility of tests carried out on such instruments.

The procedure adopted in measuring critical heights of fall was as follows. Five heights were taken in the critical region, and twenty trials were made at each height. The percentage of ignitions P was determined at each height, and the heights taken so as to provide a range 20 per cent. $< P < 80$ per cent. The P values were then plotted against H , and the value of H corresponding to $P = 50$ per cent. taken as the critical height. Not only is this the "ideal" height for the sample, but as may be readily shown by statistical theory it is the most accurate part of the experimentally obtained distribution curve.

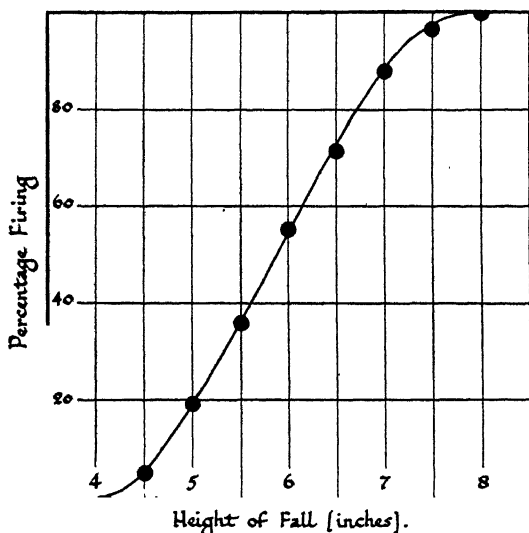


FIG. 2.—Statistical distribution of "Drop Test" ignition.

With the aid of these improved instruments and the statistical method, the conditions governing the ignition of the pellets were investigated, with the following results.

Variation of Height of Fall with Mass of Striking Weight.

It was a matter of common knowledge that the ignition of a pellet is determined both by the mass of the striking weight and also its velocity at impact, but the relationship of these two variables was not known precisely. The first experiments were therefore concerned with this point. The data tabulated in Table II are typical of all the explosives examined.

Similar experiments were carried out with the large apparatus using less sensitive explosives and the same relation found to hold when the standard steel rollers were used.

Table II.—Variation of Critical Height of Fall with Mass of Falling Weight.
Striking pins : Cylindrical steel 0·004 sq. inches cross section.

Mass of falling weight (W ounces).	Mercury fulminate.		Lead trinitroresorcinate mixture.	
	Critical height of fall (H inches).	Product (WH).	Critical height of fall (H inches).	Product (WH).
0·5	14·0	7·0	18·0	9·0
1·0	7·1	7·1	9·0	9·0
2·0	3·5	7·0	4·6	9·2
3·0	2·3	6·9	3·1	9·3
4·0	1·7	6·8	2·3	9·2

The results of all the work with the standard apparatus may be summarised by the following rule. Ignition of the sample occurred whenever the kinetic energy of the striker exceeded a critical value which was characteristic of the sample itself, that is to say, the mass m of the striker and the striking velocity v were always related by the expression

$$mgH = \frac{1}{2}mv^2 = K.$$

It follows that K may be taken as a quantitative measure of the sensitiveness of the explosive to percussion under the given conditions.

In fig. 3 (*a*), the results of Table II have been plotted in curves I and II both of which exhibit the characteristic hyperbolic form obtained from all such experiments. The test of the relation $mgH = K$ is to plot the reciprocals of m against H ; a straight line should be obtained passing through the origin and with gradient K . This has been done for the same results in fig. 3 (*b*), curves I and II.

It is curious at first sight that the striking velocity should only enter into the problem through the kinetic energy, but it appeared that for the range of velocities available this was undoubtedly the case. An extension of the work revealed the fact, however, that with other types of apparatus, the simple kinetic energy criterion did not always hold. Thus, in one set of experiments the hardened steel rollers used with the larger instrument were replaced by mild steel pistons, and not only were much greater energies required to cause ignitions, but smaller weights appeared to produce ignitions with less kinetic energies than did the larger weights. On another occasion, the pellets were held in an actual cartridge system, that is to say, at the centre of a circular

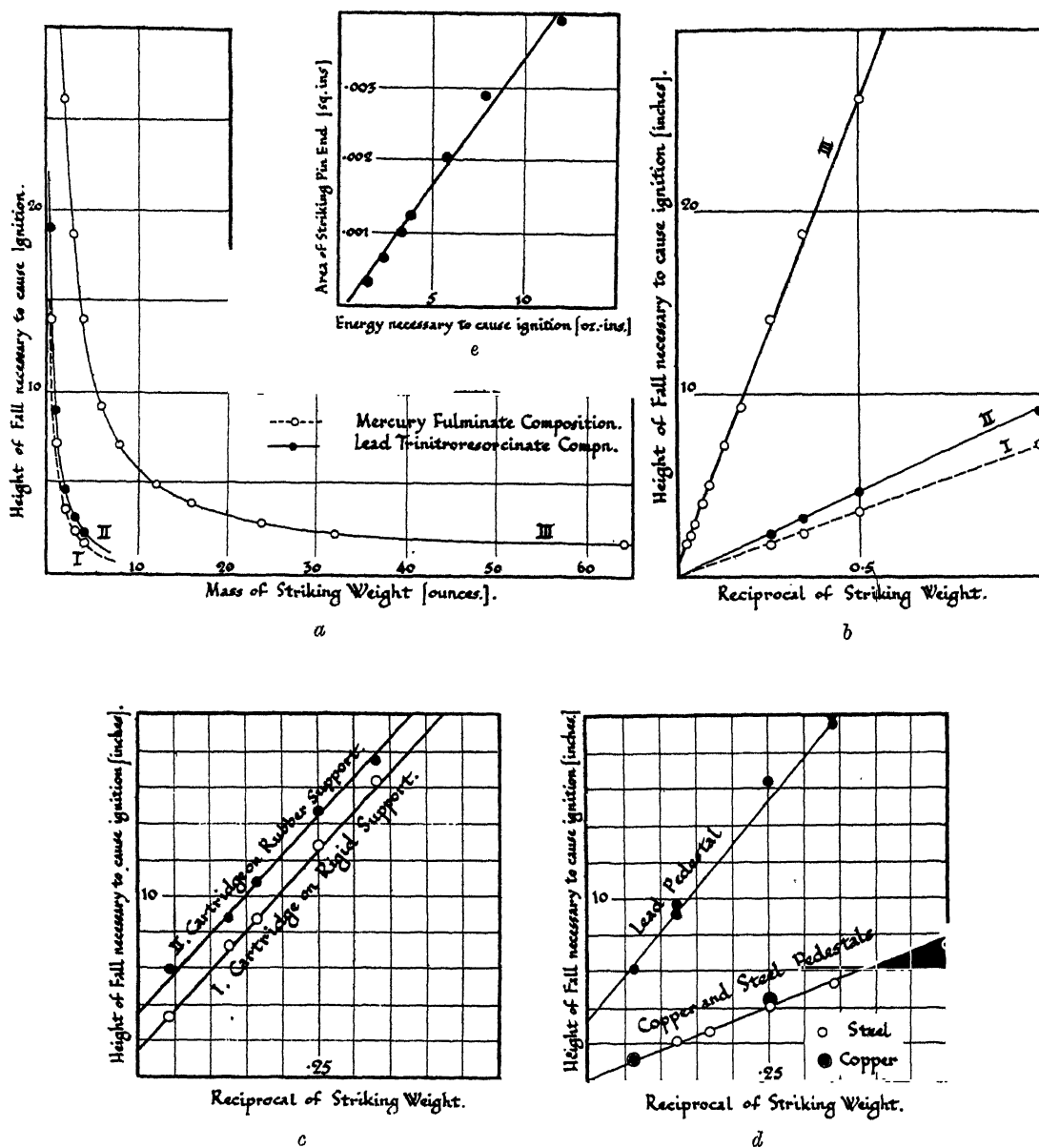


FIG. 3.—(a) : Variation of height of fall necessary to cause ignition with the mass of the striking weight. (b) : Evaluation of K , the sensitiveness of the explosive. (c) : Illustrating theory of elastic support and showing that the effect of increasing the elasticity of the support is to increase the constant H_0 . (d) : Effect of supporting pellet of explosive on various metallic pedestals. (e) : Variation of sensitiveness with area of striker end.

metal plate fixed at its periphery but capable of vibrating axially, and the results again showed a departure from the constant K criterion.* The departure from the standard case was always exhibited in the same way, *i.e.*, by a displacement of the straight line relating $1/m$ with H . As an example, the actual results with a cartridge system are shown in fig. 3 (a), curve III, and the corresponding reciprocal diagram in fig. 3 (b), curve III. The modified results were all capable of being represented by the more general formula

$$mg(H - H_0) = K.$$

An investigation of all the cases where an H_0 term was obtained revealed a factor common to them all, namely, a lack of rigidity in the supporting or striking systems. This lack of rigidity could be introduced in several ways, *e.g.*, by using a piston which was permanently deformed by the blow, by a light pedestal which could vibrate unduly, or by any lack of strength in the constraints of the pellet itself. As the effects of a yielding support are of great interest in developing a theory of propagation of detonation, this point was investigated further, starting with the simplest possible case, in which the pedestal itself was free to recoil and only offered resistance to the blow by virtue of its own inertia.

For this purpose, the pedestal with supporting system was hung as a ballistic pendulum free to recoil horizontally. The striking weight was also hung as a pendulum, and swept down an arc so as to strike the pellet in a horizontal line through the centre of gravity of the carrier pendulum. The suspensions of both pendulum systems were of the quadrifilar type. Difficulties of manipulation prevented any extensive work with this apparatus, but the results showed a perfectly definite H_0 term.

A very simple explanation may be offered in this case. If the struck system is able to recoil, then all the energy of the striking mass is not available for local dissipation at the place of contact, but only that part which is relative to the common centre of gravity, *viz.*,

$$L = \frac{1}{2}mv^2 \left(\frac{M}{M+m} \right) = mgH \left(\frac{M}{M+m} \right),$$

which may be written

$$L = mg \left(H - \frac{L}{Mg} \right).$$

* Examination of the literature showed that results of this type had also been obtained by Zelinka (Vojensko-Technické Zpravy, vol. 3, p. 111 (1926)), who attributed them to a specific effect of the striking velocity.

If the condition of ignition is that the available energy L must attain some critical value, then L may be taken as a constant and the necessary relation between m and H then becomes

$$mg(H - H_0) = \text{constant} = L,$$

where H_0 has been written for L/Mg , and this expression is then identical with the one obtained by experiment. The slope of the line obtained by plotting $1/m$ against H should be the same as in the case of a rigid support. To test this point, the results obtained with the pendulum were plotted and L found to be 6.0 ounce-inches. Another set of the same pellets were tested on the standard instrument and the K value found to be 7.5 ounce-inches. Having regard to the experimental difficulties in the pendulum work, this was regarded as sufficiently good agreement.

The problem of the cartridge system in which the yielding is due to deformation of the suspension is more complex, but the experimental results are of the same form as in the case of the pendulum. It was argued that if the appearance of an H_0 term were really due to lack of rigidity, then by introducing further freedom artificially, the H_0 intercept should be increased, but the gradient L of the line should remain the same. Two sets of experiments were carried out to test this point, and the results are given graphically in fig. 3 (c). Curve I was obtained for cartridge systems held rigidly as usual; a definite H_0 intercept is shown. The whole system was then separated from the steel pedestal by a pad of hard rubber and a new series of experiments made with the result that curve II was obtained. The anticipated effects are both present, the H_0 intercept being increased from 1.6 to 3.5 inches while the gradient remains the same. In order to obtain accurate points on these curves, the full statistical method was employed, and 1000 trials were involved in these two experiments.

Finally, experiments were made in which the pellets were supported by short pillars of various metals. Steel, copper and lead were used so as to provide different degrees of rigidity. The cylinders were $\frac{1}{2}$ inch high and carried the small brass shells in which the explosive material was pressed in the form of a thin pellet. It was not possible in this case to provide lateral constraint for the striker piston, so the alternative arrangement shown in fig. 1, c, was used, a small Hoffmann steel ball being used as the striking piston. The falling weight struck this ball vertically, and the levelling was such that neither the weight nor the ball showed appreciable horizontal motion after the impact. The lead cylinders were so deformed by the blow that a fresh one was used for each trial, the necessary standard of repeatability being ensured

by employing Standard Eley lead crushers as used for pressure-taking in firearms.

The results of this work are shown graphically in fig. 3 (*d*). Those on the steel and copper pedestals, and also those on the standard apparatus were identical, but those on the lead were very different, and not only exhibited a large H_0 term but also a greatly enhanced K value.

The special interest of these experiments lies in the fact that for those metals which were not strained beyond their elastic limits by the blow, the initiation was independent of the metal. This is important since otherwise the K values of the standard apparatus might vary with the class of steel used for the pedestal or hammer. The evidence of the experiments suggests that any metal may be used which is not permanently deformed. In illustration of this principle it may be mentioned that whereas the mild steel served quite well for experiments with the sensitive explosives, it behaved like the lead when used in the large instrument with less sensitive high explosives requiring blows of greater energy.

The work described above indicates that only when the supporting system is rigid to a high degree can the results of the falling weight test be relied upon to estimate the relative sensitiveness of explosives to percussion. If this condition is not fulfilled, then additional factors are introduced which may entirely mask the true values. The critical condition for ignition in the absence of these disturbing factors is that the energy of the falling weight must exceed a critical value, but this does not necessarily mean that the energy itself is the immediate causative agency; any other factor uniquely determined by the total energy may actually be responsible.

Effect of Area of Blow.

It seemed natural to suppose that the critical energy should vary with the area over which the blow is distributed, and this was actually found to be the case; further, the relation was strictly linear. The experiments were carried out with striking pins of different cross-section using the standard apparatus. A complete range of weights and heights was taken for each size of piston and the K values calculated in each case. The latter have been plotted against the area of the piston-end in fig. 3 (*e*), and the linearity of the relation is evident. The particular mixture used was one of mercury fulminate, potassium chlorate and antimony sulphide.

Comparison of the Critical Energy with the Amount of Explosive Initiated.

One object of using very thin pressed pellets in the above work is that the stresses developed in them should be sensibly uniform throughout their thickness, and hence also the average conditions governing the liberation of energy should not vary sensibly throughout the pellet. It is of great interest therefore to compare the energy of the striking mass with the amount of heat required to raise the explosive to its ignition temperature.

Reading from the graph of fig. 3 (e), a flat-ended piston of 0.0034 square inch required a striking energy of 10 ounce-inches for ignition of the particular mercury fulminate mixture. The pellet had a thickness of 0.0185 inch and hence the volume affected was 0.0000629 cubic inch. Supplementary experiments showed that the relation between initial thickness of pellet and the critical energy was sensibly linear. Assuming the whole of the energy of the striker to be usefully employed, in heating the pellet then the available energy amounts to 3.038×10^5 ounce-inches per cubic inch or in metric measure 68.5 joules per cubic centimetre. Converted into heat, this would be 16.3 gram calories, and assuming a density of 4 and a mean specific heat of 0.2 the temperature rise would be 20° C. approximately. The ignition temperature of this particular mixture is 163° C.

Hence, assuming the most perfect conditions theoretically possible, the energy of the striker is inadequate to heat the explosive uniformly to the ignition temperature. Actually, the efficiency of the energy transfer will be far below unity and the inadequacy becomes still more marked. Nor, under the conditions of experiment, can it be admitted that the average quantity of heat liberated in any one portion of the pellet is in excess of that liberated in another. It is true that for much less sensitive explosives, such as trinitrotoluol, very much greater striking energies are required, but an examination of the experiment leaves little doubt but that the greater part of this energy is dissipated elsewhere than in the explosive, chiefly in the steel of the pedestal and weight themselves. For example, in one series of experiments, the explosive was supported on mild steel pistons which were slightly crushed by the blow. The extent of the crushing was the same (to within less than 0.1 per cent.) whether an explosive pellet was mounted on the piston or not, although in the former case frequent ignitions were obtained. On these grounds alone, we regard many of the calculations of temperature rises in the explosive quoted in the literature as misleading.*

* For example, Muraour, 'Bull. Soc. chim. Fr.', vol. 41, p. 620 (1927) quotes as an example of the adiabatic heating effect that 5 kgm. falling 30 cm. will detonate picric acid, this

Assuming, however, that very high and localised concentrations of energy are possible in specially favoured parts of the explosive, it appeared desirable to investigate next the manner in which the sensitiveness, as measured by the K value, depended upon the initial temperature of the explosive.

Temperature Variations of Sensitiveness.

For the purpose of testing the sensitiveness of explosives at different temperatures, the standard anvil was replaced by a conical hard copper cylinder which was surrounded by an electrically heated jacket, fig. 1 (*a*). The upper end of the pillar carried a polished stainless steel top which was capable of renewal as it became pitted by the firing of the explosive. Replacements were made in practice after every 20 ignitions, otherwise the results ceased to be repeatable. The steel top was enclosed by a copper block which was drilled to serve as a guide for the striking pins, the latter being of hardened steel $\frac{3}{8}$ inch in diameter and $\frac{1}{2}$ inch long. Temperatures were measured by means of thermocouples inserted between the steel top and the copper block. Control experiments showed that the thermocouple temperature corresponded with that of the explosive. The explosive was not used in the form of a pressed pellet; instead a measured charge was placed loosely on the steel, covered with the piston, and pressed with a standard load. The heating current was then allowed to flow, and when the desired temperature was attained, the weight was released and the result observed.

On account of the relative slowness of these operations, the full statistical method, involving 100 experiments for each critical height determination, was considerably abbreviated, relying on the experience gained with many thousands of trials by the standard method. Although the K values are not quite so accurate therefore as those which have been given above, they may be relied upon to within approximately 5 per cent.

The results are given in the following table and are also shown graphically in fig. 4, the sensitive substances being mercury fulminate and guanyl-nitroso-amino-guanyl-tetrazene* ("tetrazene").

On plotting K against the temperature t , a curve is obtained which may be produced to cut the axis $K = 0$ at a value T which is obviously the temperature

energy of 1.5 kgm. metres being sufficient to raise 0.01 gm. of a substance of unit specific heat through 350° C. The point we would make is that an overwhelming proportion of this energy is incapable of becoming concentrated in the explosive, and no figures have been produced for the heat actually available.

* See British patent 201,009.

Table III.—Temperature Variation of Sensitiveness.

Explosive layer about $\frac{1}{2}$ mm. thick.

Striking weight = 2 ounces.

Mercury fulminate.			Tetrazene.		
Critical energy K ounce-inches.	Temperature, t° C.	$(T - t)$ $T = 163^{\circ}$ C.	Critical energy K ounce-inches.	Temperature, t° C.	$(T - t)$ $T = 118^{\circ}$ C.
10	157.5	5.5	20	113.0	5.0
20	149.5	13.5	30	111.0	7.0
30	140.0	23.0	40	108.0	10.0
40	119.0	44.0	50	103.2	14.8
45	87.0	73.0	55	100.0	18.0
46	37.0	126.0	60	89.0	29.0
			70	75.0	43.0
			76	15.0	103.0

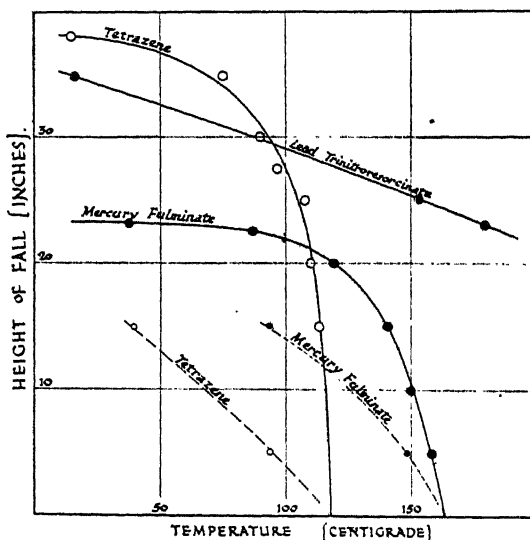


FIG. 4.—Influence of temperature upon sensitiveness to percussion.

of spontaneous ignition of the sample.* Additional columns have been added to the table giving the values of $(T - t)$, the deficiency between the actual

* If this intercept is the true spontaneous ignition temperature, then it was suggested that the use of a different striking system which would give different K values at normal temperatures, in accordance with what has been said previously concerning the effect of area of blow, would nevertheless yield curves having the same T intercept, since the ignition temperature is independent of any blow being delivered at all. The experiment was tried and the two dotted curves were obtained, confirming the above surmise in full.

and the ignition temperatures, which is compensated in some way by the energy K supplied by the striking weight.

On the hypothesis that some given fraction of K is liberated as thermal energy in some localised portion of the sample, then the rise in temperature would be very nearly proportional to K since the specific heats of the components do not change very greatly over the temperature range in question, and we should expect K to be proportional to $(T - t)$. The experimental results, however, are not even in remote agreement with such a relation. Thus, the theory of localised heating which would meet the first objection to the adiabatic compression theory, namely, the inadequacy of the available energy, leads us immediately into a further difficulty.

On account of the first objection, we have been led to reject the adiabatic compression theory and to seek in its place for a mechanism which is more in accord with the experimental evidence. The theory which is outlined below not only provides such a mechanism but includes without difficulty an explanation of the temperature effects described above. We would point out, however, that this explanation is not essential to the main argument.

Theory of the Mechanism of the Initiation Process.

It is desirable in the first place to examine the process whereby energy is liberated in the interior of the explosive, bearing in mind that the external pressures operating during the impacts range from about 10 tons per square inch in the case of mercury fulminate to over 100 tons per square inch in the less sensitive explosives, also that the explosive compositions are never rigid solids but are all capable of further compression under pressures of these high intensities.

The explosive composition contains internally a large number of free surfaces and interfaces between contiguous granules, and on the application of the external pressure, these interfaces are subjected simultaneously to normal stresses forcing the two granules into more intimate contact, and to tangential components tending to cause shearing motion. As a result of the former, strong cohesive bonding is produced by the interaction of the molecular fields, but the linkages are no sooner established than they are violently ruptured by the action of the shearing motion. The surfaces become disintegrated, the individual molecules are left in highly activated states, and the selective energisation results in the development of the phenomena of tribo-electricity and tribo-chemistry.

Both of these phenomena have the common feature that highly energised molecules are produced in numbers considerably in excess of those which would be provided by a Maxwell distribution, unless an extremely high temperature were to be postulated. The frictional heat which appears as a result of the motion is to be regarded as a consequence of the redistribution of these high and irregular energy concentrations rather than *vice versa*.

In more general terms, the continued process of surface linkage and disruption under very intense pressure results in the chemical decomposition of the medium at a velocity which is proportional to the rate at which work is being done in the process. To this decomposition must be added that due to the thermal reaction proper. If either or both of these reach a critical value, then the action becomes self-propagative, and the whole granule of explosive substance decomposes very rapidly. A third agency is then brought into play within the pellet, namely, the effect of the liberation of hot reactive gas products, and providing the other components of the mixture are chemically suitable, reaction becomes general throughout the layer, spreading very rapidly from centres provided by the elementary granule decompositions. Since the whole of the layer is still under the influence of the impact pressure, and the original conditions of excitation are in operation, the time required for the complete ignition of a properly designed composition is extremely small, and in practice the explosion appears to be simultaneous with the blow.

The theory admits of quantitative development along the following lines.

The rate of variation of the thermal reaction with temperature is governed by the well-known expression

$$k_1 = Ae^{-Q/RT},$$

in which Q is the energy increment involved in the activation process, A is a proportionality coefficient and T the absolute temperature. The other contribution to the reaction velocity is provided by the mechanical forces, and is governed by K , the energy of the blow. We may therefore set

$$k_2 = BK,$$

while bearing in mind that a more complete analysis of the process may entail using some power of K ; for example, the normal pressure and tangential stresses at a given interface are each proportional to the applied external pressure P , and for the standard apparatus P in turn is determined by the energy K of the blow and the elastic constants of the steel pedestal. Taking, however, the first power for simplicity, we have for the complete condition

$$Ae^{-Q/RT} + BK = k,$$

where k is the critical reaction velocity necessary to initiate the decomposition of the granule. This equation is therefore the quantitative expression of the relation between temperature and the sensitiveness as measured by the K value. Its great interest lies in the fact that it yields curves of precisely the same type as those given by the experimental work. If we write it in the form

$$\left(\frac{k}{B} - K\right) = \frac{A}{B} e^{-Q/RT},$$

and put $K_0 = k/B$, then

$$\log (K_0 - K) = \log \left(\frac{A}{B}\right) + \frac{Q}{RT}.$$

As $T \rightarrow 0$, $K \rightarrow K_0$ so that the experimental curves should be asymptotic to the same line $K = K_0$, and this tendency is well marked.

Secondly, if the reciprocals of the absolute temperature are plotted against the Napierian logarithms of the differences between the K values and the asymptote of the curve, the result should be a straight line. This also was found to be the case.

The slope of these lines should be equal to Q/R and hence the critical increment of the assumed activation process may be estimated, subject to the provision mentioned above that K and not some power of K is the appropriate variable to be plotted. For mercury fulminate, the calculation gave $Q = 16.8$ kilogram-calories per gram mol, whereas had \sqrt{K} been selected as the variable, the result would have been $Q = 26.5$. As the Q value is by hypothesis characteristic of the crystal of mercury fulminate, it seemed of great interest to obtain an independent estimate of the critical increment of the purely thermal decomposition. Professor Garner very kindly undertook this investigation at our request,* and obtained the figure 29.4 kilogram-calories per gram mol.

The tentative theory outlined above is in accordance with the following facts, which have been established quantitatively by means of the two standard instruments but which, in the interests of brevity, will not be quoted here in detail.

(1) The results of covering the granules with oil or wax reduces the sensitiveness enormously and this finds a ready explanation on the tribo-chemical

* Result privately communicated: we understand that a description of this work is in course of publication.

hypothesis, since the seat of the initiation is presumed to be the interfaces of the granules.*

(2) A soft sensitive material may be sensitised by the admixture of harder components, whether inert or not.

(3) The grist of the granules plays a very important part. The smaller the particles the closer is the packing and the less opportunity is there for breakdown and relative motion. Thus as the granules are made very fine, the percussion sensitiveness falls, and may even cause the pellet to become inert.

(4) If opportunity is given for increased relative motion, as, for example, by striking a glancing blow, the sensitiveness increases rapidly. It may be remarked that such few tribo-chemical reactions as have been studied are associated with grinding forces and not direct impacts.

(5) If the pre-compression of the pellet is increased greatly a decrease in the sensitiveness is observed as the density increases. On the theory developed above, this is only due to the very much restricted opportunity afforded for internal motion in a mass in which the voids have been reduced to negligible dimensions.

(6) The specific chemical action of certain non-explosive compounds in modifying the K value of a mixture may easily be understood. Such substances are generally oxidising agents like the perchlorates, peroxides, and nitrates. If we consider an interface between an explosive which is almost invariably oxygen-deficient, and one of these granules, then while the decomposition of the former is proceeding, the products react with the oxygen carrier and give rise to further products of high energy content. The result is to bring a third agency into play increasing the reaction velocity at the surface of the explosive, and supplementing the frictional effects. It is now known that products formed by highly exothermic reactions at a surface are powerful catalysts, and when they are concentrated at a reactive surface under a pressure of many tons to the square inch it is not surprising that they exert a great influence.

* Dr. N. K. Adam, with whom we discussed this question, very kindly furnished us with a series of fatty acids, viz., nonylic, lauric, palmitic and eicosoic acids in which the lengths of the carbon chains were progressively lengthened, and we hoped to obtain a corresponding series of sensitiveness values when they were used in turn to cover the crystal surfaces of mercury fulminate. Unfortunately the effects were quite masked by the fact that under the falling weight, the crystals were ruptured and initiation commenced at the fresh unlubricated surfaces thus opened up.

Application to the Theory of Detonation.

The experiments described in this paper have been concerned with the specific sensitiveness of thin layers only. It remains therefore to indicate briefly how work along these lines may further our knowledge of the phenomena of the propagation of detonation.

At the outset it may be remarked that similar sensitiveness experiments may be carried out equally well with gunpowder and other typical non-detonating compositions. The ignitions are not usually so violent as those obtained with sensitive detonating explosives; thus the ignition of 0.4 grain of mercury fulminate sometimes shattered the hardened steel piston, and pitted a steel requiring a critical deformation pressure of over 380 tons per square inch. Again, charges of 0.2 gram of gelatine were often initiated with sufficient violence to throw a 2 kilogram hammer 10 cm. in the air. It is evident that only the most rapid and complete decomposition of the whole sample could produce these results.

With these facts in mind, let us consider the process of initiation and propagation of detonation in a cartridge of high explosive. For simplicity, suppose the initiation to be due to the shock wave from a neighbouring explosion. The aerial pulse traverses the gap with a velocity considerably greater than that of sound, and impinges upon the first layer of the explosive. The necessary criterion for the initiation of the first layer is that the stresses set up across it must be equivalent to those which would be given to the layer if placed on a rigid support and struck by a rigid hammer with a blow of energy K per unit area. (We may refer to these as the K conditions.) For example, in the case of certain dynamites, the stress in the layer would have to exceed 80 tons per square inch, and as the cartridge is mechanically equivalent to a yielding support, the pressure in the air pulse would require to be much greater than this figure. There is usually no difficulty in this; the pressure exerted by the complete explosion of mercury fulminate, lying uncovered on a steel plate, certainly exceeds 400 tons per square inch and the reason for the great initiating power of fulminate will readily be understood.

In the absence of any reaction in the first layer, the pressure in the wave-front would be damped very rapidly indeed in a medium possessing the physical properties of a typical high explosive. The limited amount of energy would be rapidly dissipated as heat, owing to the internal motion, and the pulse would quickly become degraded to an ordinary sound wave. (We exclude in this connection the case of propagation in a liquid.)

In the case of detonating explosive, however, the first layer will be initiated by the applied stress, and the hot gaseous products of reaction will contribute to the local pressure, compensating for the loss of energy suffered by the pulse in traversing the layer. There is no lack of potential chemical energy available for this purpose, as a rule; thus a woodmeal dynamite containing 60 per cent. nitroglycerine has an available chemical energy of 950 gram-calories per gram.

It is not sufficient, however, to have this energy in the potential form; it must be liberated at a sufficiently high velocity to be able to reinforce the pressure in the wave-front. If we consider a layer of the same thickness as that used in the percussion experiments, namely, 0.02 inch, then the pressure wave travelling at the slowest possible velocity met with in normal detonations, *i.e.*, about 2000 metres per second, would traverse the layer in 2.5×10^{-7} second, and therefore the decomposition must in this space of time provide sufficient pressure on its own account to compensate for the loss otherwise suffered by the wave in travelling the 0.02 inch. Failing this, the velocity of the wave will fall, and continue to do so until either (1) it is sufficiently slow to permit the reaction velocity to maintain the pressure, or (2) the pressure in the wave-front falls below the critical value necessary to produce the K conditions. In case (1) the detonation proceeds at its characteristic stable velocity, while (2) the detonation comes to an end and the undetonated cartridge is merely hurled forward by the gas pressure.

In certain cases, the energy liberated by the initiating pulse is more than sufficient to compensate for the loss suffered by the original wave, and the velocity of the wave-front is then increased in sympathy with the higher driving pressure behind it, so that the detonation wave accelerates to the characteristic velocity.

The characteristic velocity of detonation of a high explosive arises therefore as a result of the balance of two qualities inherent in the nature of the explosive, namely, (a) the specific sensitiveness which determines the critical pressure conditions necessary for the initiation of each layer, and (b) the velocity of chemical reaction in a layer once initiated under the influence of the critical pressure conditions. In dynamic equilibrium at the wave-front, the latter must maintain the pressures required by the former, in other words, there must be a stable velocity of propagation if the explosive is to detonate at all.

If we were to assume a simple adiabatic heating effect, then (a) could be computed from a knowledge of the specific heats and compressibility data,

but the conclusion arrived at as a result of the experimental work is that (a) depends upon the tribo-chemical potentialities of the medium, and is affected by a large number of factors which must be determined by experiment. The other quality (b) of the medium, depends upon the chemical nature of the composition and also to some degree upon the physical conditions, and it is here that the reason is to be found why many explosive mixtures like gunpowder which give reasonably low K values in the fall-hammer test do not detonate. Although initiated by tribo-chemical action, the reactions are too slow, even under the high pressure, to reinforce the pressure wave in a short enough space of time. Hence the necessity for the presence of some constituent which is capable of extremely rapid and explosive decomposition once initiated. Into this class of substances fall the pure explosive compounds such as mercury fulminate, lead azide, tetryl, picric acid and nitroglycerine.

The precise mechanism of this rapid decomposition offers another field of investigation of great theoretical interest, though not necessary for the present theory, in which it is sufficient to accept the experimental facts.

The above brief sketch of a tribo-chemical theory of detonation must be regarded as a fruitful hypothesis based upon the experimental results described in the paper, as well as being in accord with a mass of information on the subject of detonating explosives which cannot be enlarged upon here. The chief conclusion is that this branch of physical chemistry is essentially an experimental one, and that its first requirement is the accumulation of precise data. The specific sensitiveness of an explosive, and the reaction velocities under high pressures are not entirely amenable to theoretical calculation; the underlying physical and chemical principles are simple, but in the first instance are too interwoven for other than experimental treatment. At the same time, they involve in a very interesting manner, fields of theoretical chemistry which are of especial interest at the moment, namely, the study of the mechanism of surface reactions, of molecular force fields, and of molecular activation, and indicate the practical importance of what is at present the rather obscure and neglected subject of tribo-chemistry.

The above work was carried out in the Research Department, Nobel Section, Imperial Chemical Industries, Ltd., at Stevenston. We are indebted to Messrs. Imperial Chemical Industries, Ltd., and in particular to Mr. T. Donaldson and Dr. J. Weir for permission to publish this paper.

Summary.

An attempt has been made to approach the general investigation of the phenomena of detonation in solid explosives by a study of the dynamics of the initiation of thin layers of explosives by impulsive forces. The forces are most conveniently applied by controlled impacts of masses of steel, as in the so-called "drop-weight" or "fall-hammer" tests.

It has been shown that whereas the results of such experiments exhibit large fluctuations, they can be made to furnish very reliable data by the application of statistical methods, and carefully controlled experimental conditions.

A number of quantitative relations have thereby been obtained relating the criterion of ignition with the mechanical and geometrical factors involved in the impact.

It is shown that ignitions may be caused by impacts having insufficient energy to raise the sample of explosive to the ignition temperature, and hence some more detailed theory of the ignition process is necessary.

The temperature variation of the ignition process has been studied and an exponential relation found, which is difficult to reconcile with a simple theory of local temperature rise, but suggestive of a surface activation mechanism; some possibilities of the latter are indicated tentatively.

The natural extension of the work as a basis for the formulation of the general mechanism of detonation is discussed.

Mobility of Alkali Ions in Gases.

By C. F. POWELL and LUANG BRATA, The Wills Physical Laboratory, University of Bristol.

(Communicated by A. P. Chattock, F.R.S.—Received May 26, 1932.)

1. Introduction.

In a recent paper* an account was given of the determination of the mobilities of ions of the four alkali metals, sodium, potassium, rubidium and caesium in the gases helium, neon and argon. In this paper the work in these gases is completed by the addition of the results for lithium ions and the determinations are extended to the gases krypton, xenon, nitrogen and hydrogen. In addition the four-gauze method is applied to :—

- (a) An investigation of the clustering of impurity molecules round the simple alkali ions.
- (b) An attempt to detect the two isotopes of lithium by the difference in their mobility.
- (c) A search for the missing element No. 87.

For these investigations it was necessary to increase the efficiency of the method in separating ions of slightly different mobilities, and, before discussing the results, we shall analyse briefly the factors which determine this resolving power.

2. Resolving Power.

In the analysis it is convenient to make use of optical analogies. If the charge moving through a gas is being carried by several groups of ions of different mobility, then it is possible by means of the four-gauze method to find how the charge is distributed amongst the various groups. A curve drawn from any particular experiment, showing the relative charge carried by the different groups plotted against the mobility, may be called the mobility spectrum of the ions in the particular gas used. Similarly we may speak of the "resolving power" of the method, and by this we mean its capacity to separate ions of different mobility. We may define the resolving power as the peak frequency divided by the frequency range of the peak, at half maximum, due to a single group of ions. In the previous paper curves were given showing the mobility of the alkali ions plotted against the mass of the ions for the three lightest rare gases. We may refer to these as the "dispersion curves" for

* Tyndall and Powell, 'Proc. Roy. Soc.,' A, vol. 136, p. 145 (1932).

the three gases. It was shown that in the case of argon the dispersion curve for the alkali ions was represented with considerable accuracy by the expression

$$k = B(1 + m/M)^{\frac{1}{2}}, \quad (1)$$

where k is the mobility of the ion of mass, M , moving in the gas of atomic weight m . Though in general the results for a particular gas may not be expressed by such a simple relation, we may expect the dispersion to increase with the molecular weight of the gas and this has been shown to be the case in the experiments with krypton and xenon. With an apparatus of given resolving power one could therefore expect to separate two ions of different mass more easily in the heavier gases. The resolving power itself can be increased by applying the following considerations.

As in the four-gauze method, let us introduce into an electric field a pulse of ions of a given kind in the form of a thin "plate" and consider its thickness after it has passed through the field. (1) It will have what we may term an instrumental width determined by the original thickness of the pulse. This in practice cannot be made indefinitely small, a limit being set by the sensitivity of the electrometer employed to measure the ionic current. (2) The pulse will broaden as it moves through the field owing to diffusion and self-repulsion. This effect will be smaller, the stronger the field in which the ions move. Due to (2) the pulse may be said to have a natural width as distinct from its instrumental width due to (1). Expressed in terms of a mobility spectrum a given spectral line will have a certain width, part of which is instrumental and part natural.

It is evident that in order to obtain good resolving power it is necessary to arrange that the ion pulses generated by the first pair of gauzes shall be thin and that the ions shall move in high electric fields so that the broadening due to diffusion may be small. The relative importance of the factors (1) and (2) in limiting the resolving power is shown in the three curves (a), (b) and (c) of fig. 1. The peaks are all due to caesium ions moving in argon at a pressure of 10 mm. of mercury. In curve (a) the values of the alternating potential applied to the grids B and D, fig. 2, was 9.5 volts and the main field, 67 volts/cm. In curve (b) all the potentials have been doubled so that the ions pass through the apparatus twice as quickly as before and any broadening due to diffusion and self-repulsion is reduced. The resolving power has been increased from 9.5 to 11.7 as a result. To reduce the instrumental width curve (c) was taken with the main field increased from 133 to 145 volts per centimetre. This is practically equivalent to a reduction in the value of the alternating potential

keeping the main field constant. At this higher main field the peak frequency is increased and the width of the pulse of ions at the new peak frequency is smaller than that at the old. The resolving power has now increased to 14.8. This resolving power is sufficient for most purposes and it is much higher than that obtained by any method recorded in the literature.

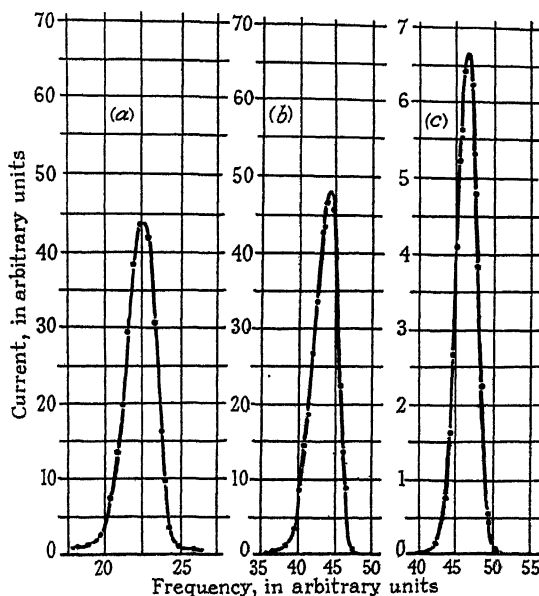
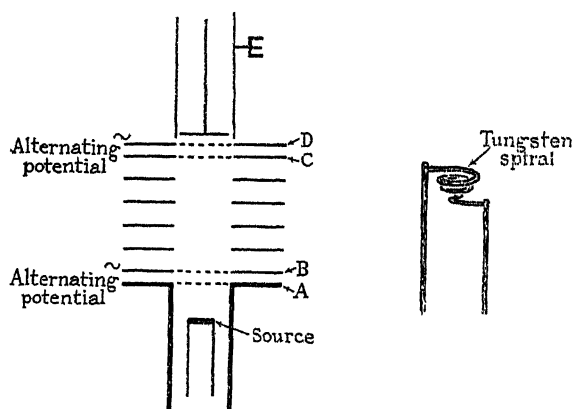


FIG. 1.—Curves showing successively increasing resolving power. Cæsium ions in argon. Argon pressure = 10.06 mm. of mercury. (a) Main field (BC), 67 volts/cm.; alternating potentials on B and D, 9.5 volts. (b) Main field (BC), 133 volts/cm.; alternating potentials on B and D, 19 volts. (c) Main field (BC), 145 volts/cm.; alternating potentials on B and D, 19 volts.

3. Experimental Technique.

The experimental tube is similar to that described previously and the metal parts are shown diagrammatically in fig. 2, *a*. The steady values of the potentials between the plates were kept constant at the following values: AB = -2 volts, BC = 100 volts, CD = -2 volts, DE = 8 volts. Alternating potentials of 9.5 volts were applied to the plates B and D. With these values of the potentials, and with the alternating potentials out of phase, the peaks corresponding to a given single type of ion appear in five orders when the usual current-frequency curve is determined. A typical curve for sodium ions moving in argon is shown in fig. 3. The peaks in the higher orders are also plotted on a greater current scale.

With these experimental conditions, assuming that the diffusion coefficient is proportional to the mobility, the peak due to a single kind of ion in a given

FIG. 2, *a*.FIG. 2, *b*.

order should have a given form independent of the pressure and the mass and size of the ions, any departure from this form being due to the fact that the

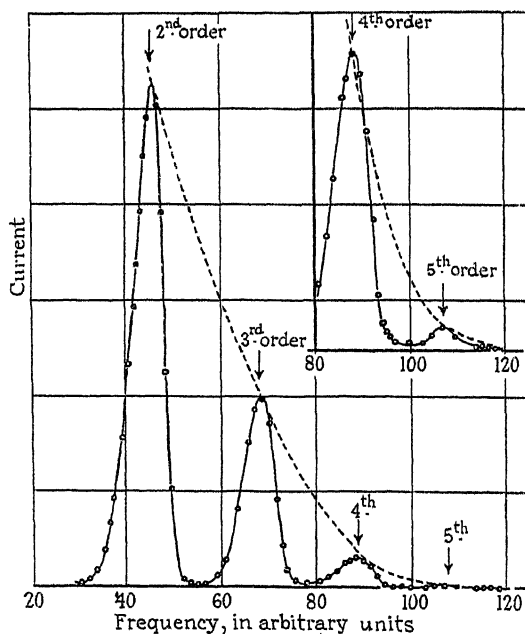


FIG. 3.—Sodium ions in argon. Curves showing peaks in 2nd, 3rd, 4th and 5th orders.

ions are not all of the same kind. If the peak due to a single kind of ion, in a gas at a pressure of p millimetres of mercury, occurs at a frequency f , then the

quantity fp is proportional to the mobility of the ions. fp is known for helium ions moving in helium gas, using the third order peak with the alternating potentials out of phase. The corresponding mobility is 21.4 cm./sec./volts/cm. The mobility corresponding to any other ion in any other gas can thus be found.

In the previous work no difficulties were experienced due to "overlapping spectra." In gases giving higher dispersion and also in examining the effects of impurities, we have found many examples of complications due to this effect. The resulting curves can be analysed from our knowledge of:—

- (a) The total current in any given order corresponding to a given emission from the source.
- (b) The relative intensities of the peaks due to a given kind of ion in the different orders.

An example of overlapping is shown in fig. 8.

We may also consider the effect due to a change in the nature of the ion by a collision process in the main field; the ion may, for example, pick up a neutral impurity atom and cluster, or it may transfer its charge to an atom of lower ionisation potential. In general these collision processes lead to a reduction in the mobility of the ions. Suppose that ions of type A of high mobility all change half way across the main field to ions of type B of lower mobility. The time they take to cross the main field will correspond to a mobility between that of ions A and ions B, and a peak will appear corresponding to this intermediate value. In a given order this peak will be of very much smaller intensity than if the charge had entered the main field either as ions of type A or type B and had retained their character. This is so because the ions arriving at the second shutter grids are of type B and in order to reach the collecting electrode, they must cross the shutter at a higher frequency than that corresponding to B ions in the given order we are considering. The number of ions crossing the shutters falls off very rapidly with increasing frequency, as is shown by the dotted line in fig. 3, and the intensity of the peak due to the ions which have suffered change in the main field will be correspondingly reduced. Accordingly it is found experimentally that very little disturbance is produced by changes in the nature of the ions in the main field.

4. Dispersion Curves for Alkali Ions in Krypton and Xenon.

The experimental values for the mobility of the alkali ions in krypton and xenon are shown in Table I. In each column two values are given corresponding to the mobility of a given ion in a given gas. That on the left is the

Table I.—Mobility of the Alkali Ions in Krypton and Xenon.
(Values in cm./sec./volt/cm.)

Gas.	Li ⁺ .		Na ⁺ .		K ⁺ .		Rb ⁺ .		Cs ⁺ .	
	Expt.	Theo.	Expt.	Theo.	Expt.	Theo.	Expt.	Theo.	Expt.	Theo.
Krypton	4.03	4.03	2.34	2.42	1.98	1.98	1.61	1.58	1.44	1.47
Xenon	—	3.22	—	1.86	1.50	1.50	1.12	1.19	0.99	1.01

experimental value; on the right the value deduced from an equation similar to (1) making the experimental and theoretical values for a given gas to coincide for potassium ions. The accuracy of the measurements was limited by the determination of the pressure. The gas pressure was of the order of a few millimetres and the pressure measurements were made on a wide-limbed mercury U-tube manometer. The deviations of the experimental from the theoretical values are of the same order as the experimental error.

The alkali ions were obtained from Kunsman sources prepared by the method previously described and heated in a tungsten spiral of the type shown in fig. 2 (b). When the lithium source was first heated the emission consisted almost entirely of sodium and potassium ions, but after several hours these diminished in intensity and lithium ions began to come off. Eventually, after continuous heating for about 12 hours, lithium ions alone were emitted. The advantage of using a spiral containing a comparatively large amount of the Kunsman mixture is that one obtains a source of very long life. It has the additional advantage, especially in working with sources containing metals of high ionisation potential, that the necessary high temperature can be obtained without fear of burning out the filament. The ordinary method of spreading a thin layer of the mixture on a platinum filament is unsatisfactory in this respect when using lithium and the alkaline earths, and the mixture sometimes drops off the platinum.

5. *Lithium Ions in the Rare Gases.*

In order to complete the series of measurements with the alkali ions in argon, neon and helium, the mobility of lithium ions in these gases was also determined and the values obtained are shown in Table II, together with those for the alkali ions previously obtained (Tyndall and Powell, *loc. cit.*). Using these values, the complete dispersion curves are shown in fig. 4, the mobility being plotted logarithmically.

Table II.—Mobility of Alkali Ions in Gases.
(Values in cm./sec./volt/cm. at 760 mm. of mercury, 20° C.)

Gas.	Li ⁺ .	Na ⁺ .	K ⁺ .	Rb ⁺ .	Cs ⁺ .
Helium	25.3	23.2	22.4	21.0	19.0
Neon	14.35	9.00	8.00	7.18	6.58
Argon	4.99	3.22	2.78	2.39	2.24
Krypton	4.03	2.34	1.98	1.61	1.44
Xenon	—	—	1.50	1.12	0.99
Hydrogen	13.3	13.6	13.5	13.4	13.4
Nitrogen	4.21	3.04	2.70	2.39	2.25

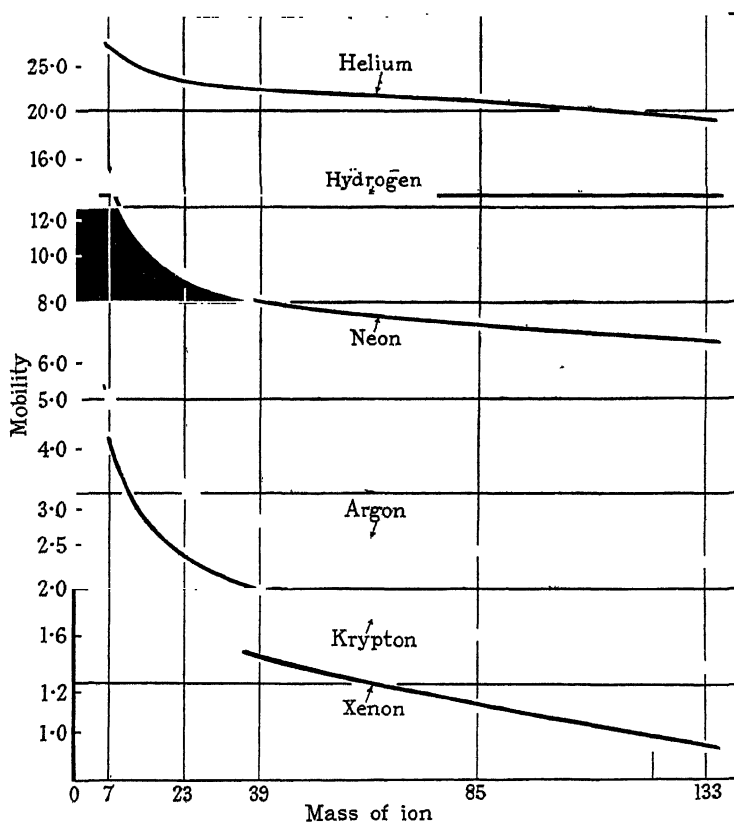


FIG. 4.—Dispersion curves for the rare gases and hydrogen.

If the abundance ratio of the isotopes $\text{Li}_6 : \text{Li}_7$ is 1 : 12 the peaks due to lithium ions moving in the heavier rare gases should show some evidence for the presence of Li_6 ions. Since the experimental results in argon are represented with considerable accuracy by equation (1) we should expect the ratio of the

mobility of Li_6^+ ions to Li_7^+ ions to be 1.07 : 1 in this particular gas. The curve obtained when two peaks of the type shown in fig. 1 (c) of intensity ratio 12 : 1 and mobility ratio 1 : 1.07 are compounded is shown in fig. 5. In working with lithium sources in argon and krypton we have found evidence of a broadened curve of this type. This suggests that a concentration of Li_6 atoms might be produced by making use of the difference in the mobility of

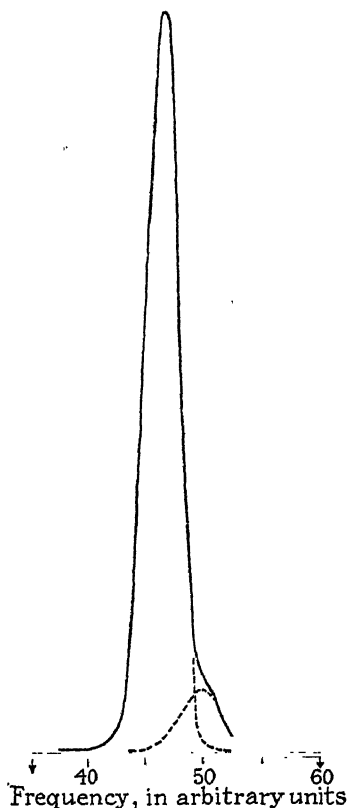


FIG. 5.

the two isotopes. Using a Kunsman source in conjunction with the Franck modification of the Rutherford method, the charge first reaching the electro-meter plate, as the frequency of the alternations applied to the plates is reduced, will consist largely of ions of the lighter isotope. The method would have the advantage over positive ray methods that theoretically the plates could be of indefinite extent. The chief difficulty would seem to be in producing a suitable source of ions. Tyndall and Powell (*loc. cit.*) showed that a Kunsman source has the disadvantage that it evaporates large numbers of neutral atoms which could condense on the surface collecting the ions.

6. Comparison of Results in Argon, Krypton and Xenon.

In order to estimate the mobility of ions of mass M moving in a gas with molecules of mass m , Langevin assumed that the ions and molecules might be regarded as elastic spheres between which there was an attractive force,

varying inversely as the fifth power of the distance and due to the polarisation of the molecules by the ions. He obtained the following expression for the mobility, k , in terms of the density of the gas, ρ , and the dielectric constant, D ,

$$k = \frac{A}{\sqrt{\rho(D-1)}} \left[1 + \frac{m}{M} \right]^{\frac{1}{2}}, \quad (2)$$

where the quantity A is a function of $(D-1)$ and of the distance of closest approach, σ , of the ion and the molecule in collision. If we consider the

mobility of different ions in the same gas, in general we should find therefore that the mobility would depend both on the mass and the size of the ions. This is found to be the case in neon and helium (Tyndall and Powell, *loc. cit.*), for example, where the quantity $k/(1 + m/M)^{\frac{1}{2}}$ decreases as we go from Li^+ to Cs^+ .

A simplification occurs, however, in the case of highly polarisable gases for which the quantity A in equation (2) tends the limiting value 0.51 if k is measured in centimetres per second per unit electrostatic field. The physical significance of this limiting value is that the polarisation forces between the ions and the gas atoms are large at distances greater than σ so that the ion can, in the majority of its collisions, receive large deflection without "colliding" with the gas atom. In such a gas, equation (2) reduces to equation (1) which has been shown to represent the experimental results for the alkali ions in argon, krypton and xenon with considerable accuracy. Using the values of the dielectric constant of the rare gases given by Watson* we have deduced the value of A necessary to fit the experimental facts represented by the dispersion curves in these gases. The results are shown below :—

Table III.

Gas.	A.
Langevin's theory	0.51
Argon	0.54
Krypton	0.55
Xenon	0.56

In view of the assumptions made in Langevin's theory the agreement of the experimental values between themselves and with Langevin's value is remarkable.

It would be more satisfactory to replace the elastic sphere model by that of point centres of repulsive and attractive forces as has been done by Hasse.† The constant A in Langevin's formula is then replaced by one which varies with the relative magnitude of the repulsive and attractive force coefficients. But full data for the application of this model to collisions between the various alkali ions and rare gas atoms are not yet available. In the limit, however, when the gas is highly polarisable the value of A approaches that given by

* 'Proc. Roy. Soc.,' A, vol. 132, p. 569 (1931).

† 'Phil. Mag.,' vol. 12, p. 554 (1931); vol. 1, p. 139 (1926).

Langevin and the two models become virtually identical. The agreement between the theoretical and experimental values suggests that in most of the collisions of the ions with the atoms the distance of closest approach is such that the force due to polarisation predominates and any failure of the model chosen to express the real forces between the reacting particles at closer distances becomes unimportant.

7. *Experiments in Nitrogen and Hydrogen.*

Values for the mobility of the alkali ions in nitrogen and hydrogen are shown in Table II and the dispersion curve for nitrogen in fig. 6. If the alkali ions

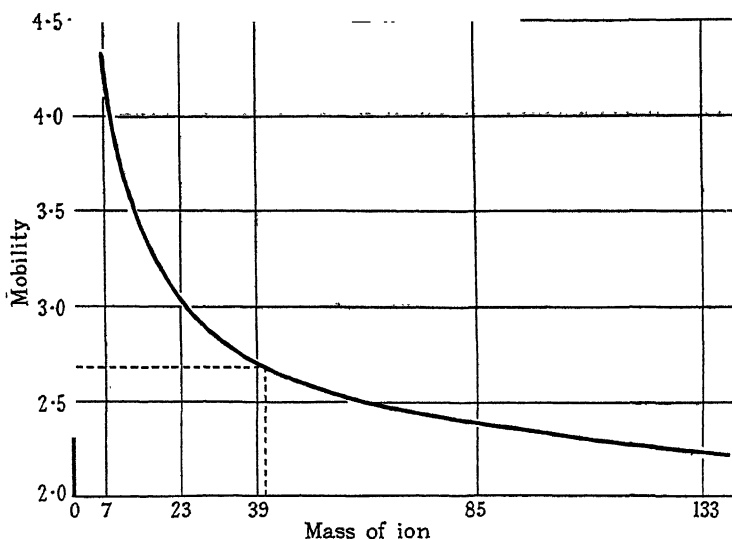


FIG. 6.—Dispersion curve for nitrogen.

in hydrogen followed the simple mass law, the difference in mobility between lithium ions and caesium would be 13 per cent., most of which occurs between lithium and sodium. But it will be seen that in hydrogen the mobility is independent of mass over the whole range and equal to 13.4. At first sight this suggests that all the original alkali ions had undergone an increase in mass by the addition of foreign molecules before entering the main field. In the usual method of preparing a Kunsman source it is finally sensitised by reduction in the apparatus by heating in hydrogen. In working with such a source in hydrogen we might, therefore, expect a continuous production of water molecules unless the source had been previously very completely reduced.

We have carefully investigated the possibility that these results are due to clustered ions, since Loeb* has reported the presence of a sodium ion of mobility 17 in hydrogen. The following are some typical results using a sodium source. The source was first reduced and the apparatus evacuated. Pure hydrogen was then admitted and the source heated at the temperature necessary to give the required emission. Liquid air was placed on a trap near the apparatus. Any water produced by further reduction of the source will diffuse continuously to this trap and there will be an equilibrium concentration of water molecules in the path of the ions. The chance of an ion capturing a water molecule will depend upon :

- (a) The equilibrium concentration of water molecules. This will decrease with decreasing hydrogen pressure owing to the greater rate of diffusion of the water to the trap. The concentration will be still further reduced if the rate of production of water is also lowered by the reduction of pressure.
- (b) The time taken by the ions to cross the apparatus. Given a fixed field, this is proportional to the pressure.

For these reasons we may expect any clustering effect due to collisions between sodium ions and water to decrease very rapidly with decreasing pressure. This is found to be so experimentally. In fig. 7, curve (a) was taken in hydrogen at 14.5 mm., curve (b) at 11.8 mm. and curve (c) at 8.7 mm., the fields remaining constant throughout the series. The results all support the view that the peak at mobility 13.2 is due to an unclustered sodium ion. The change in mobility from 11.2 to 13.2 is due to the addition of a single water molecule to the faster ion ; otherwise we should obtain evidence for ions of intermediate mobility. If the faster ion were in reality Na^+ , H_2O then the curve (c) taken at the lowest pressure should show some evidence for the true Na^+ unless the attachment of the second water molecule takes place very much less easily than the first. Actually we have worked at considerably lower pressures than that at which curve (c) was taken ; the mobility of the ions remains constant until the value of E/p becomes so high that the speed is no longer proportional to it.

The interpretation of the hydrogen results is complicated by the fact that the dispersive power of the gas for ions of different mass is zero within the limits of experimental error. The lowering of the mobility by the addition of water must therefore be due to an increase in the " size " of the ion. The formation

* 'Phys. Rev.,' vol. 38, p. 549 (1931).

of clusters can be much more easily followed in nitrogen, for which the dispersion curve for alkali ions is shown in fig. 6, or in the heavier rare gases. Except for the value for lithium the results in nitrogen can be represented by an equation similar to equation (1) which suggests that in this gas also the size of the ion is of little importance. The value of A of equation (2) which fits the experimental values for nitrogen is 0.48. This is lower than the value for argon. Although there can be little chance of water being present in appreciable

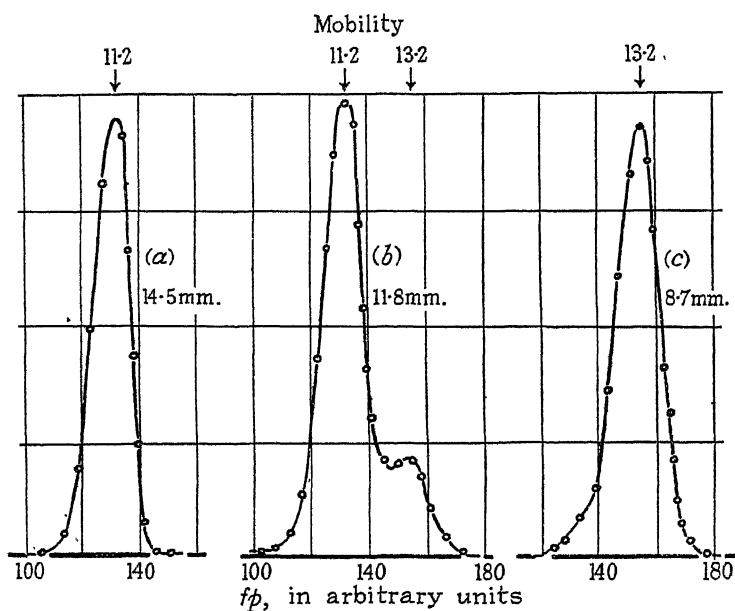


FIG. 7.

quantities when working with a Kunsman source in pure nitrogen in a rigorously degassed apparatus, it was found experimentally that clustered ions were frequently obtained at the higher pressures. As in hydrogen, all trace of these clustered ions can be removed by working at sufficiently low pressures. As the pressure is increased, using a sodium Kunsman source, the first clustered ion to appear has a mobility of 2.66. From the dispersion curve in fig. 6 it will be seen that the mass of an "alkali-like" ion with this mobility would be 41. Since any increase in the size of the alkali ions apart from an alteration in mass must, if it produces any change, lead to a decrease in the mobility of the ions, we can state that the mass of this clustered ion must be 41 or less, so that the mass of the clustering molecule must be 18 or less. The only two polar impurities which could lead to this result are water and ammonia, one molecule in each case.

It was pointed out above that it is unlikely that water could have been present. The following test shows definitely that the clustering in nitrogen is not due to water. In two experiments, one in hydrogen and the other in nitrogen, the pressures of the two gases were adjusted so that the time taken by sodium ions to cross the apparatus was the same in the two cases. In these conditions the times taken by water molecules, generated at the source, to reach the liquid air traps must be nearly equal. It was found that the proportion of clustered ions was considerable in nitrogen and zero in hydrogen. The rate of production of water cannot be greater in nitrogen than in hydrogen and the clustered ions cannot therefore be due to this impurity. On the other hand, if the nitrogen contained traces of hydrogen, ammonia would be present since the Kunsman source is an ammonia catalyst. Further, the vapour pressure of ammonia is considerable at liquid air temperatures and this would explain the comparative difficulty in avoiding clustered ions in nitrogen. We conclude that the clustered ions in nitrogen are due to the addition of ammonia molecules to the original alkali ions.

At the higher pressures the attachment of a second and a third molecule to the original clustered ion was observed. The clustered ion Na^+ , 2NH_3 was found to have a mobility corresponding to an alkali-like ion of the same mass. This is further evidence that in a highly polarisable gas, between wide limits, the size of the ion has little influence on its mobility.

For a complete examination of the clustering of ions it would be more satisfactory to add a known concentration of molecules of a polar impurity to a rare gas, rather than to work in nitrogen where the impurity control is difficult. In this way it would be possible to determine the average number of gas kinetic collisions between an alkali ion and molecules of polar impurity before attachment takes place. An experiment is now being undertaken with an apparatus in which the ions pass through a variable distance in a constant field before being analysed. From results obtained with the present apparatus it seems that a water molecule attaches itself more readily to the lighter alkali ions than to the heavier and, further, that the maximum number of water molecules which can attach themselves to the ion decreases with increasing atomic weight.

8. *A Search for Eka Cesium in Samarskite.*

Bainbridge* has used the Dempster method to make a search for the presence of the missing alkali, element No. 87, in various minerals. He obtained

* Bainbridge, 'Phys. Rev.', vol. 34, p. 752 (1929).

negative results. His method consisted in examining the positive rays emitted from a Kunsman source containing the alkali extracted from these minerals. Recently Papish* has reported the presence of this alkali in the mineral samarskite. Although the present method is very much less sensitive than the positive ray method in detecting a trace of one alkali in the presence of others, we have used it to search for eka caesium since the alkali content of samarskite, and similar minerals which are possible sources of the element, is very variable. Further, the concentration of eka caesium in the alkali extract, if it is present at all, can probably be increased by chemical means since the chemical properties of its compounds can be predicted with some confidence.

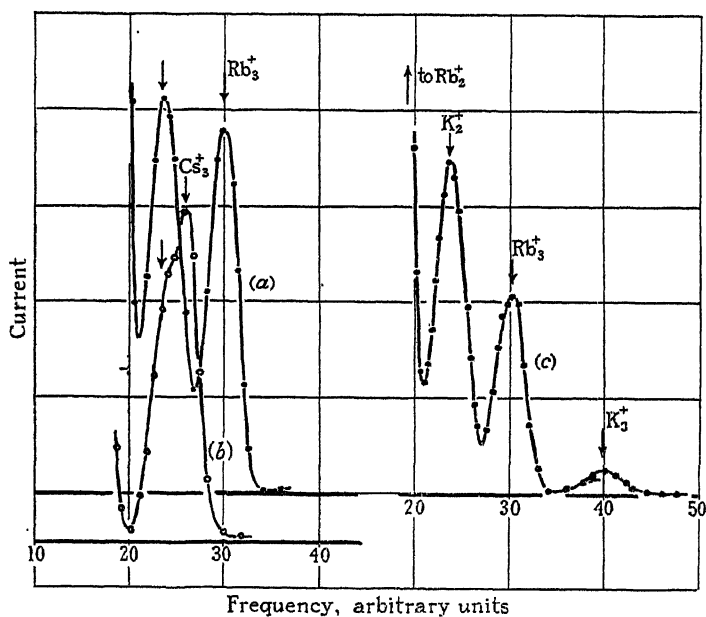


FIG. 8.

We are indebted to Mr. J. G. Pilley for supplying us with the alkali extracted from 2 kilograms of Norwegian samarskite. The least soluble fractions of the platino-chloride of this alkali were used in a Kunsman source. This source was mounted in a tube together with a source of caesium and a source of rubidium ions. The caesium and rubidium sources were first reduced in hydrogen and finally the one containing the samarskite alkali. The apparatus was filled with xenon and the ions emitted by the caesium and rubidium sources analysed. The resulting curves are shown in fig. 8, curves (a) and (b). The

* 'J. Amer. Chem. Soc.,' vol. 53, p. 3818 (1931).

curves show that two types of ions are being emitted by each source. The peak occurring at the lower frequency has the same mobility in the two cases ; the higher frequency peaks are due to rubidium and caesium respectively. If it is assumed that the lower mobility peak is a third order peak like those of rubidium and caesium, the mobility, and hence the mass, of the ions giving rise to it can be determined since the dispersion curve for xenon can be expressed by a relation similar to equation (1). The mass so found is 220, which is in good agreement with the atomic weight to be expected for eka caesium. The peak may, however, be due to a trace of ions of higher mobility in a lower order, and a second run with the rubidium source showed this to be the case. Curve (c) in fig. 8 shows the presence of a small peak due to the presence of potassium ions in the third order. A large proportion at least of the low frequency peak in curves (a) and (b) must be due therefore to potassium ions in the second order. When the samarskite alkali itself was heated it was found to emit potassium ions in large quantities without any measurable trace of caesium and rubidium. Evidently the presence of potassium in the ions from the rubidium and caesium sources was due to the evaporation of the potassium from the samarskite alkali source during the reduction process. This phenomenon has been described before (Tyndall and Powell, *loc. cit.*) and we have obtained many examples of it.

Owing to the large percentage of potassium in the samarskite source, xenon is not a convenient gas with which to disperse the ions. The third order peaks due to any eka caesium ions which may be present is completely masked by the second order peak due to potassium ions. We therefore examined the ions from the source in argon where the dispersion is not so great, and where the peak due to eka caesium would occur at a greater frequency than the second order potassium peak. No trace of eka caesium ions was found, and if present their concentration must have been less than 1 per cent. In a Kunsman source containing a mixture of alkalis the ions of the heavier elements are emitted highly preferentially owing to their low ionisation potential. If we accept Bainbridge's estimate that the ionisation potential of eka caesium is less than that of caesium, we can conclude that the concentration of the eka caesium in the concentrated samarskite alkali is less than one in four thousand.

In conclusion, we wish to acknowledge our indebtedness to Professor A. M. Tyndall, for his interest and for many discussions throughout the course of this work. One of us, L. B., is indebted to the Rockefeller Foundation for a fellowship which enabled him to pursue this work.

Summary.

(1) The determination of the mobility of ions of the alkali metals in the rare gases has been completed by measurements in the gases krypton and xenon, and the results have been extended to include lithium ions.

(2) The results in argon, krypton and xenon are expressed to within a few per cent. by the relation,

$$k = \frac{0.55}{\sqrt{\rho(D-1)}} \left[1 + \frac{m}{M} \right]^{\frac{1}{2}} \text{cm./sec./unit e.s. field,}$$

where ρ is the density of the gas at N.T.P., D is its dielectric constant, m is the mass of the gas atom and M the mass of the ion. If we replace the factor 0.55 by a term A this relation is identical with two formulæ, one deduced by Langevin on an elastic sphere model and the other by Hassé on a point centre of force model. When polarisation forces predominate A tends to a limiting value 0.51 in both cases.

(3) Similar experiments have been made in nitrogen and hydrogen. When traces of polar impurities are present the original alkali ions become clustered. The ions formed by the addition of one, two and three molecules of ammonia can be identified by their mobility, and a quantitative study of the growth of clustered ions becomes possible.

(4) The four-gauze method has been applied to search for the alkali eka caesium in the mineral samarskite, with negative results.

The Distribution of Suspended Particles Under Gravity.

By C. M. McDOWELL and F. L. USHER, The University, Leeds.

(Communicated by R. Whytlaw Gray, F.R.S.—Received May 27, 1932.)

[PLATE 4.]

The apparent stability and freedom from settling shown in practice by many colloidal suspensions has in recent years prompted renewed study of the distribution of the particles of colloidal systems under the influence of gravity, and results greatly at variance with earlier work have been obtained. The object of the present communication is (1) to show that the later conclusions are not justified by the experiments on which they are based, and (2) to describe experiments which favour a considerable extension of the applicability to colloidal solutions of a law similar to that which defines the behaviour of ordinary dilute solutions.

Perrin,* in 1909, showed that there was, within a depth of 0.01 cm. from the surface, a logarithmic decrease of concentration with height for suspensions of gamboge or mastic in gravitational equilibrium. In 1914 Costantin,† working at lower depths than Perrin, and with a higher volume concentration of gamboge and correspondingly higher numerical concentration, obtained evidence of a repelling force between the particles at concentrations greater than 8×10^{10} particles per cubic centimetre. Perrin showed how his equation could be modified to take this into account by assuming a law of the van der Waals' type, which was found to agree with the observations up to the highest concentration investigated (6×10^{11} particles per cubic centimetre). In the following year Westgren‡ published the results of a series of measurements on selenium and gold sols with coarse particles, which were in complete harmony with Perrin's law over a distance of 0.05 cm. and at concentrations up to 5×10^{10} per cubic centimetre. These investigators were precluded from studying the distribution over greater ranges of depth solely by the restrictions of their experimental method, which demanded that for accurate counting the number of particles in the field of view should not be too large. For this reason also they were limited to moderately dilute sols and to rather large particles, that is, to specially prepared sols.

* 'Ann. Chim. Phys.,' vol. 18, p. 1 (1909).

† 'C. R. Acad. Sci. Paris,' vol. 158, pp. 1171, 1341 (1914).

‡ 'Z. phys. Chem.,' vol. 89, p. 63 (1915); 'Z. anorg. Chem.,' vol. 93, p. 231 (1915).

Burton,* in 1921, first studied the sedimentation of "ordinary colloidal solutions," and his work, in common with later investigations, indicates that, whatever the condition of the surface layer examined by Perrin may be, the concentration at lower depths and in the bulk of the liquid is uniform, and it is claimed that extreme departures from "ideal" behaviour occur at particle concentrations even smaller than those studied by Perrin. The evidence for uniformity of concentration with depth is either the behaviour of sols under ordinary laboratory conditions which in fact prohibit the attainment of equilibrium, or is provided by experiments which, as will be shown, are open to serious objections.

Burton and Bishop† attacked the problem of distribution equilibrium by means of an experimental arrangement entirely different from Perrin's, viz., by studying sedimentation in large vessels. When the sols had been allowed to stand in tall tubes for a long time samples were withdrawn at various levels and analysed, and a uniform concentration was observed over the whole tube length.

Porter and Hedges‡ extended Burton's work by verifying his statement that concentration became uniform with increasing depth, and also found the law of change between that region and the upper one in which the laws of dilute solutions are valid. Rejecting Burton's explanation of the uniformity as a consequence of electrical repulsion between the particles, Porter and Hedges examined the applicability of a formula derived from a form of van der Waals' equation from which the " a " term was omitted, that is, the observed deviation from simple behaviour was treated as a volume effect. Their curve had a point of inflection at $n = n_{\infty}/3$ leading to a uniform concentration, $n_{\infty} = 1/b$, at large values of h . They evaluated their constants by fitting the theoretical to the experimental curve, and from one of the constants (K) thus obtained the radius of the particles was calculated. A microscopic method was employed for determining concentration. The gamboge sol used was obtained undisperse by fractional centrifuging, but the radius of the particles was not determined.

Barkas§ attempted to make Porter and Hedges' equation less arbitrary and thus remove one objection to it, by determining a relationship between its constants and the nature and dimensions of the colloidal particles used. The

* "Physical Properties of Colloidal Solutions," 1921.

† 'Proc. Roy. Soc.,' A, vol. 100, p. 414 (1922).

‡ 'Trans. Faraday Soc.,' vol. 18, p. 91 (1922); vol. 19, p. 1 (1923).

§ 'Trans. Faraday Soc.,' vol. 21, p. 66 (1925).

radius of the particles in a unidisperse sol was deduced from one of the constants (K) and also determined practically by applying Stokes' law to centrifuging data. Fairly good agreement was found with copper, but with gamboge and silver it was not so satisfactory.

Data relating to the conditions observed in experiments on distribution are summarised in Table I.

Table I.

Author.	Material.	Radius of particles in cm.	Depth of cell in cm.	Time required for equilibrium calculated.	Time allowed for equilibrium.	Highest concentration observed (particles per c.c.).
*Perrin	Gamboge	1.2×10^{-5}	0.012	80-160 mins.	3 hours. No change after 5 hours	Not stated
† „	„	2.1×10^{-5}	0.010	„	3 hours. No change after 15 days	Not stated
‡Costantin	„	3.3×10^{-5}	0.10	7.5-15 hrs.	3-4 days	6×10^{11}
§Westgren	Selenium	5.8×10^{-6}	0.05	1.5-3 days	Up to 7 days	2.8×10^{10}
„	Gold	6.6 to 3.3×10^{-6}	0.05	6-12 hrs.	„	4.7×10^{10}
Burton and Currie	Copper	Not unidisperse	94	?	50 days	Not stated
¶Porter and Hedges	Gamboge	Not determined	0.5 ?	?	Not stated	10^7
** „	Paraffin	$ca. 2 \times 10^{-4}$	Not stated	?	„	10^6
††Barkas	Gamboge	1.5 to 3.0×10^{-5}	1.2	10-20 days	„	3.2×10^9
„	Silver	2.6×10^{-6}	1.2	„	„	3.0×10^8
„	“Copper”	?	1.2	?	„	3.3×10^8

* 'C. R. Acad. Sci. Paris,' vol. 146, p. 967 (1908). † *Loc. cit.* ‡ *Loc. cit.* § *Loc. cit.* || 'Phil. Mag., vol. 47, p. 721 (1924). ¶ *Loc. cit.* (1922). ** *Loc. cit.* (1923). †† *Loc. cit.*

It is evident that two points have to be considered. There is first the striking discrepancy between different sets of experimental results, illustrated by the work of Porter and Hedges, and of Barkas on the one hand, who have found deviations from the laws of dilute solutions at particle concentrations of 10^6 per cubic centimetre, with a uniform concentration at 10^7 per cubic centimetre; and that of Perrin, Costantin, and Westgren on the other, who have found normal behaviour up to concentrations greater than 10^{10} particles per cubic centimetre. There is also the question of the interpretation of Burton's observations and of the later experiments on distribution. It would be unreasonable to suppose that the simple exponential formula can hold good in regions where the suspended particles themselves occupy a considerable fraction of the total volume; the question is whether the laws of dilute solutions break down completely at very low concentrations. The results of

the later work indicate that they do ; for example, in one of the silver sols studied by Barkas a uniform concentration was observed at 2.55×10^7 particles per cubic centimetre or about 3×10^{-9} actual volume concentration, figures which for any true solution represent an almost ideal degree of dilution.

Several explanations put forward to account for such abnormal behaviour have been criticised by Kraemer,* and the only one which need be discussed here is that advanced by Porter and Hedges, who considered that the uniform concentrations observed were determined simply by the "co-volume," analogous to the " b " of van der Waals. This was calculated for their gamboge suspension to be more than five million times the actual volume of the solid, that is, the value of b deduced from their experimental curve was five million times the volume calculated from the constant K , so that the suspended particles apparently occupied the whole volume of the suspension. Recognising the difficulty raised by this conclusion, Porter and Hedges considered that the large b really represented spheres of mutual repulsion. It may be noted that if (as was done) b is deduced from a limiting value of the particle concentration it cannot represent an attraction or repulsion as signified by van der Waals' intrinsic pressure term, since the latter, whatever its sign or value, does not give rise to a limiting concentration. It follows that b is either an "effective volume," or is devoid of physical significance.

A further possibility lies in a consideration of the osmotic pressure exerted by the ions which form the outer atmosphere of the suspended particles and are electrostatically connected to them. This matter will be dealt with in a later communication, and here it is only necessary to say that the effect, calculated for the case of ordinary sols which have not been dialysed against specially purified water, would probably not be detectable below a concentration of 10^{13} to 10^{14} per cubic centimetre, and therefore need not be considered at present. One is thus forced to conclude either that the recorded deviations are due to an "effective volume" which may amount (as in a silver sol studied by Barkas) to more than 10^8 times the volume of the particles themselves, or that the deviations do not really exist.

With regard to actual experiments, an essential condition, without which no degree of accuracy in determining the particle concentration is of any value, is that of ensuring that the systems studied are actually in a state of equilibrium when the concentrations are measured. It is calculated† that the time required

* 'Colloidal Symposium Monographs,' vol. 5, p. 81 (1927).

† Mason and Weaver, 'Phys. Rev.,' vol. 23, p. 412 (1924); Weaver, *ibid.*, vol. 27, p. 499 (1926).

to attain equilibrium is between once and twice the time required for a particle to fall from the highest to the lowest level in the liquid, and, during the whole of this period adequate precautions must have been used to prevent mixing due to convection or other causes. It is in this respect particularly that the experiments of Burton and Currie, Porter and Hedges, and Barkas are at a disadvantage compared with those of Perrin, Costantin, and Westgren, which were conducted on a scale so small that the fulfilment of the requisite conditions presented no special difficulty.*

Detailed criticism of the former group is difficult because relevant data are incompletely recorded. Burton and Bishop certainly left their sols for a time long enough for some effect to be apparent, but temperature gradients in the experimental vessels and temperature changes in the surroundings were ignored. In a later series of experiments Burton and Currie used a crude (unregulated) thermostat in which the tube was immersed "almost to the top," and in which the greatest observed temperature variation was 7°. The experiments of Porter and Hedges were done at "room temperature," but it was not stated how and within what limits the temperature of their cell was controlled, nor was the time allowed for the attainment of equilibrium mentioned. According to Weaver's† calculation gamboge particles of 1.6×10^{-5} cm. radius, in a cell 5 mm. deep,‡ would require 125–250 hours for equilibrium. Barkas increased the depth of the cell to about 1.2 cm. (not counting the inlet tube shown in his diagram) and thereby increased the time required for equilibrium to between 240 and 480 hours, besides magnifying the risk of serious trouble from convection currents. He stated that the latter were a source of great difficulty and that their causes were never entirely eliminated; and further, that on account of the tendency to drifting "it was seldom possible to work for more than 2 hours consecutively." The agreement found between values of the radius of the particles calculated from centrifuging experiments and those derived from the distribution curves was satisfactory only for the copper sols. Even this, however, may be regarded as fortuitous, in view of the probability that the "copper" examined by Barkas was a complex copper-

* Convection currents in a liquid are subject to the laws governing viscous flow, and their velocity is therefore greatly reduced when one of the dimensions of the containing vessel is very small. Moreover, for a given temperature gradient in the surroundings, the temperature differences causing convection are smaller in a small vessel than in a large one.

† *Loc. cit.*

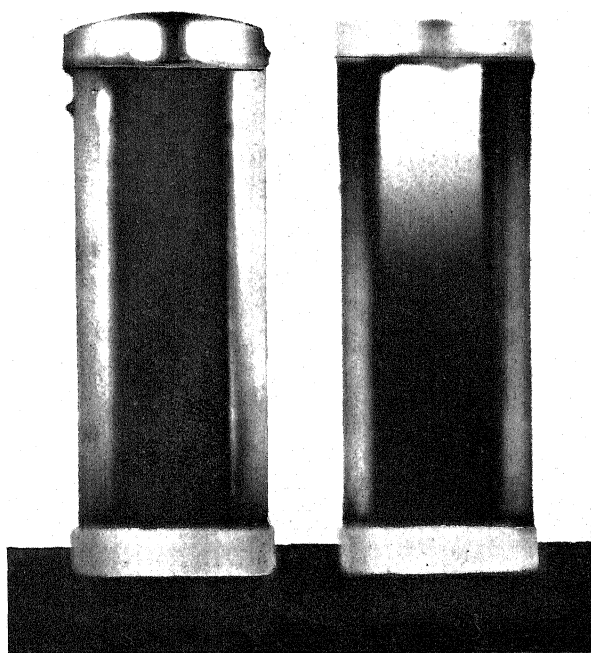
‡ Porter and Hedges, *loc. cit.*, p. 98. The cell illustrated is 5 mm. deep, and it is here assumed that this or a similar cell was used in the distribution experiment.

pyridine compound. We have found that the brown solid obtained by coagulating the colloidal "copper" solution prepared by Pieroni's method* (used by Barkas) with a neutral electrolyte, after being thoroughly washed with water and alcohol and dried over calcium chloride, contains less than 20 per cent. of copper, and evolves pyridine on heating. The density was slightly lower than that of bromoform (2.89). It was not further examined, but it is evident that an agreement found by calculating with an effective density of 8.2 cannot be maintained when that quantity assumes the value 1.89. The existence of this discrepancy lends additional support to the view that the conclusions reached cannot be considered as established without further experiment.

Experimental.

Qualitative.—Experiments similar to those of Burton and his co-workers have been carried out with multidisperse sols of silver and gold under varied conditions. When gold sols of different degrees of dispersion were examined without special precautions to ensure constancy of temperature, the results appeared to lend support to the existence of a "critical radius," since coarse (blue or purple) sols showed visible sedimentation when allowed to remain undisturbed, whereas finely dispersed red sols showed no apparent change over long periods. A similar difference was observed with silver sols. When, however, experiments of this kind were repeated under carefully controlled temperature conditions the distinction between coarse and fine sols disappeared. For example, a red gold sol made by Zsigmondy's formaldehyde method was enclosed in a glass tube 4 cm. long and of 7 mm. bore, which was sealed and suspended in water contained in a large silvered cylindrical vacuum vessel. After a few days a slight diminution of colour at the top and an intensification at the bottom was observed, and after several weeks the bulk of the liquid appeared nearly colourless, with a deep red layer occupying 2-3 mm. at the bottom. On withdrawing the tube carefully, mixing began almost immediately, and was complete after a time which varied in different experiments. A similar result was obtained when the tube was kept in the centre of a 1-lb. roll of cotton wool. These observations show that the fine sols acquire a non-uniform distribution under the action of gravity only when they are allowed to remain for a long enough time in sufficiently small temperature gradients. This is illustrated in fig. 1, Plate 4, which shows a photograph of two rectangular cells, 2 cm. high, each filled with a sample of the same silver

* 'Gazz. Chim. ital.,' vol. 1, p. 197 (1913).



A

B

FIG. 1.

sol. One, A, had been kept for 15 days in a cupboard in which the greatest variation of temperature was about 7° , as in the experiment of Burton and Currie; the other, B, had stood for the same time in a thermostat of special construction in which the temperature variation did not exceed 0.001° per hour. At the end of the time the upper half of B appeared quite colourless, whilst there was a marked gradation of colour in the lower half. The radius of the particles which would fall through 1 cm. in 15 days is about 1.9×10^{-6} cm., or about half the "critical radius" as given by Barkas. The contents of B gradually became uniform again when the cell was taken out of the thermostat, A and B being finally identical in appearance. Subsequent observations of A during a period of 9 months failed to detect any sign of settling.

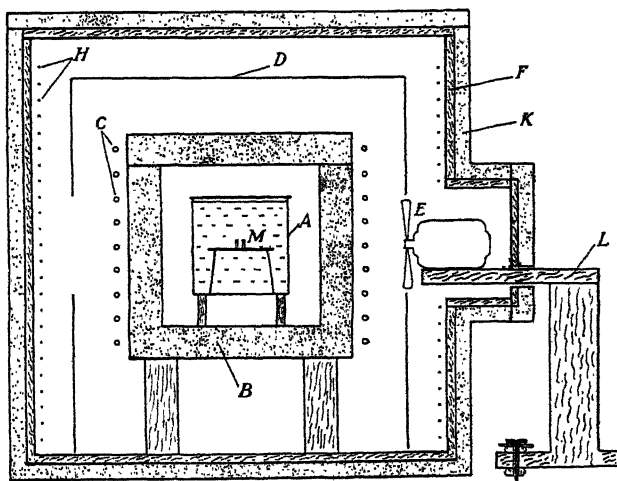


FIG. 2.

Quantitative.—The distribution of particles over a range of several millimetres, and at concentrations greater than those hitherto studied, has been determined as follows. Unidisperse gold sols, derived from a "nuclear" sol by reduction with hydrazine, were allowed to remain for 3–4 weeks in an enclosure designed to avoid the production of excessive temperature gradients. The arrangement is shown diagrammatically in fig. 2. When tested with a Beckmann thermometer under conditions less favourable than those of the actual experiments, the variation of temperature at the centre of the enclosure rarely exceeded 0.001° per hour, and was usually undetectable. This result was obtained by placing the cells, or thermometer bulb, at the centre of a large mass (3 kg.) of water contained in a tank, A, with plane parallel glass windows, which was itself surrounded by, but thermally insulated from, a space under

thermostatic control. The insulation consisted of a cubical box with double walls the 2-inch space between which, B, was tightly packed with goose down. On the outside of this box was wound a coil of glass tubing, C, containing the benzene used to operate the temperature regulator. Surrounding the box carrying the regulator was a large cubical partition, D, in two opposite sides of which circular openings were cut. One of these openings accommodated the blades of a fan, E, which was kept running continuously, whilst the other discharged the air blown by the fan into an outer space enclosed by a large wooden box, F. The inside of the latter was wound with nichrome wire, H, which was heated by current supplied through a relay operated by the temperature regulator. The outside of the box was heavily lagged, K. Suitable plugs and windows were provided, and these, together with the cubical partition, were removed when it was desired to photograph the cells. In order to avoid mechanical disturbance the fan motor was supported by a wooden bracket, L, which was secured to the table outside through rubber cushions, and was only in contact with the thermostat box through a loose flannel packing. Uniform illumination, at right angles to the plane of the diagram, was secured by allowing light from a Pointolite lamp to fall on the ground surface of a large piece of opaque white glass, the scattered light from which then illuminated a plate of finely ground glass placed behind the tank containing the cells. The positions of the source of light and of the white glass were adjusted so that a plate exposed without the cells in position showed uniform darkening over the whole field. When photographing the cells the best results, with respect to gradation of density, were obtained with "ordinary" rather than panchromatic plates. Regularity of development was secured by continuous brushing.

The cells, M, containing the sols were right-angled isosceles wedges made from a selected borosilicate Crown glass which had been proved to cause no change in the sols over a period of several weeks. These wedges were completely immersed in the water in the inner tank with one of their equal sides uppermost and horizontal, and at the end of an experiment were photographed side by side with a "standard" wedge. The standard wedge contained either (a) a diluted specimen of the same sol as was used in the experimental wedges, in which case a direct comparison was possible, or (b) a coloured substance (cobalt nitrate solution or a gelatin jelly of the gold sol) having optical properties different from those of the plain sol. In the first case the standard wedge was introduced, with elaborate precautions against thermal disturbance, at the end of an experiment; in the second it was in position from the start, since in this case the colour distribution would remain unaffected by gravity.

Measurement of the concentration of gold at different levels was based on the proportionality between absorption of light and the total quantity of gold in the layer traversed, which has been found to hold in the case of red gold sols containing particles with radius less than 3×10^{-6} cm. The problem consisted in finding the levels in a wedge of homogeneous* sol which transmitted light producing the same effect on a photographic plate as was produced by light from the same source traversing certain measured levels in a wedge containing the sol after distribution had occurred; the total quantity of gold in optically equivalent layers being the same, the relative concentration could be obtained by dividing this quantity by the thickness of the layer in question. The experimental wedges were used standing on their 45° angles, in which position the thickness of the absorbing layer decreased as the concentration of gold increased, in order to secure a partial compensation for the very great differences in absorption which otherwise would have occurred. When a standard wedge containing a substance other than a specimen of the same sol as was undergoing distribution was used, this was first photographed with the experimental cell before distribution, and their optically equivalent layers were determined. Then at the end of the experiment the levels in the experimental wedge after distribution could be expressed in terms of levels in the same wedge before distribution, that is, in terms of actual concentration, through the calibrated standard, which was the same in both photographs.

The determination of optically equivalent layers was made by mounting strips of the negative, containing the images of the separate wedges, on a carrier attached to a steel bar the far end of which could be moved parallel with the direction of a vertical side of a wedge by means of a micrometer screw. The image of a straight lamp filament was focussed on the strip, and the light, after traversing this, was received on the cathode of a vacuum photoelectric cell, the current from which was measured after suitable amplification. All the strips in turn were thus examined at measured intervals, and curves showing the variation of photoelectric current with distance along the wedges were drawn. By thus comparing photographs of the experimental and standard wedges only when they were on the same plate, the effect of any variation in exposure or development was eliminated. A Moll thermopile was used in some cases as an alternative to the photoelectric cell.

At the end of an experiment the upper (broad) part of the experimental

* *Homogeneous* is used to denote uniform concentration of gold; *undisperse* to denote equality of particle size.

wedge was often so faintly coloured that the concentration could not be satisfactorily measured in this region; thus the measurements which were designed to include the whole height, 1 cm., of the liquid, actually extend over a range varying from 3 to 8.5 mm., and to a depth of 9 mm. The highest concentration measured was 10^{12} particles per cubic centimetre.

The gold sols were prepared from a "nuclear" sol, the particle concentration of which was found, from 21 observations of the rate of fall of particles in coarse sols grown from it, to be $0.90 \pm 0.02 \times 10^{13}$ per cubic centimetre. The limits of size of the particles used in the distribution experiments were approximately fixed, a lower one because the time required for equilibrium to be established over a depth of 1 cm. could not be conveniently, or even safely, prolonged beyond about 4 weeks, the longest time for which the thermostat fan could be depended upon to run continuously without renewal of the brushes. An upper limit was set by the necessity of avoiding too great a difference in colour between the top and bottom of the cell. The results, given in Tables II and III, are shown graphically in fig. 3, in which the logarithms of the particle concentrations have been plotted against depth. This relation must be linear over any range in which Perrin's law holds good. For comparison, the results of Porter and Hedges and of Barkas, and those for one of Westgren's gold sols, have been plotted on the same scale. It should be noted that the values of $\log n$ are actual values, whereas the h values have been increased or diminished by arbitrary amounts in order to avoid overlapping. Actual depths are given in the tables.

The eight graphs have been calculated by the method of least squares, and relate to four distribution experiments, the results of which have been treated

Table II.—Sol I.

A.		C.		B.		D.	
Depth in mm.	$\log n$.	Depth in mm.	$\log n$.	Depth in mm.	$\log n$.	Depth in mm.	$\log n$.
4.44	10.36	2	9.53	2	9.76	1	9.48
5.06	10.51	3	9.82	3	10.03	2	9.73
5.67	10.63	4	10.23	4	10.34	3	10.04
6.30	10.75	5	10.50	5	10.57	4	10.36
6.90	10.89	6	10.76	6	10.77	5	10.62
7.53	11.05	7	11.02	7	11.05	6	10.84
8.15	11.22	8	11.31	8	11.30	7	11.05
8.65	11.39	9	11.66	9	11.64	8	11.31
						9	11.64

Table III.—Sol II.

E.		G.		F.		H.	
Depth in mm.	log n.	Depth in mm.	log n.	Depth in mm.	log n.	Depth in mm.	log n.
5.6	11.12	4.0	10.48	5.8	11.04	4.0	10.38
5.9	11.22	4.5	10.68	6.2	11.17	4.5	10.55
6.2	11.31	5.0	10.83	6.6	11.29	5.0	10.74
6.5	11.40	5.5	10.97	7.0	11.41	5.5	10.89
6.8	11.49	6.0	11.10	7.4	11.52	6.0	11.04
7.2	11.58	6.5	11.23	7.8	11.64	6.5	11.23
7.6	11.69	7.0	11.37	8.2	11.77	7.0	11.41
8.1	11.81	7.5	11.53	8.6	11.91	7.5	11.57
8.6	11.97	8.0	11.70			8.0	11.76
		8.5	11.89			8.5	11.94

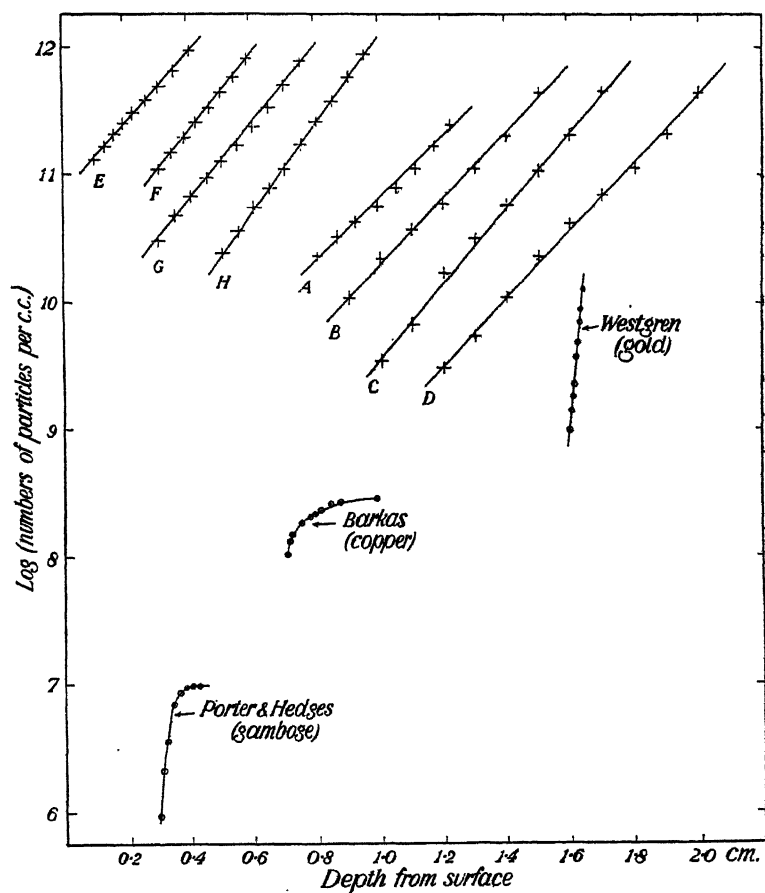


FIG. 3.

in different ways in order to define their limits of trustworthiness. A, B, C and D refer to a sol made with 6 c.c., and E, F, G and H to one with 7 c.c. of the stock solution of chlorauric acid to the same quantity of "nuclear" sol. The particulars are given in Table IV.

Table IV.

Sol.	Time allowed for equilibrium in days.	Standard of comparison.	Method of comparing densities (P = photoelectric cell, T = thermopile).
A	22	Gold gelatin jelly	P
B	29	"	T
C	(Duplicate of B, photographed at the same time.)	"	
D	29	Gold sol	T
E	23	Cobalt nitrate	P
F	27	"	P
G	27	Gold sol	P
H	(Same photograph as G.)	"	T

If Perrin's law is followed the graphs of A, B, C and D should be parallel straight lines, whilst those of the others should also be parallel but sloping rather more steeply on account of the larger particle size. Any departure from "ideal" behaviour would make the lines curve towards the h -axis as the concentration increased. It is evident that the former condition is nearly fulfilled, and that there is no tendency for the lines to bend over at the higher concentrations. The slope of each line ($d \log n/dh$) is equal to $wN/2 \cdot 3026RT$ (w = effective weight of a particle), from which the radius can be calculated. The results are given in Table V.

Table V.

Graph.	r from distribution.	r from independent data.
	cm.	cm.
A	1.44×10^{-6} cm.	
B	1.47×10^{-6} "	
C	1.53×10^{-6} "	
D	1.48×10^{-6} "	
Mean	$1.48 \pm 0.01 \times 10^{-6}$ cm.	$1.48 \pm 0.01 \times 10^{-6}$ cm.
E	1.50×10^{-6} cm.	
F	1.55×10^{-6} "	
G	1.55×10^{-6} "	
H	1.61×10^{-6} "	
Mean	$1.55 \pm 0.02 \times 10^{-6}$ cm.	$1.56 \pm 0.01 \times 10^{-6}$ cm.

The most important sources of error are (1) departure from unidispersity in the sol, (2) insufficient time for equilibrium, (3) convection currents due to inadequate temperature control, (4) irregular illumination or development, and (5) errors in photometry. The effect of (1) is to make the graphs convex to the h -axis, the convexity being mainly restricted to the bottom of the curves with finer, and to the top with coarser, admixtures. No measurements made within 1 mm. of the top or bottom have been included. The presence of (2) or (3) would lead to the curves being similar to those of Porter and Hedges and of Barkas. The calculated time required for equilibrium in the two sols is 13 to 26 and $11\frac{1}{2}$ to 23 days respectively, and the upper limit was exceeded in both cases. A comparison of the graphs for the same sol photographed again after an interval of several days indicates that equilibrium had actually been attained. There is no evidence that the distribution was seriously affected by convection currents in the sols. Both (4) and (5) might affect the results in various ways, and the only means of deciding to what extent these errors operate is to compare the results of different measurements of the same photographs, and of different photographs of the same cell. This has been done in the case of G and H, and B and D, respectively. The observed differences are probably determined almost entirely by errors of this kind.

The curves obtained by Porter and Hedges and by Barkas are of a type entirely different from that given either by the earlier experiments of Perrin, Costantin, and Westgren, or by those now described, and on the evidence available it can hardly be doubted that the difference is due to the combined effects, in the former case, of want of adequate temperature control and of too short a time being allowed for the attainment of equilibrium. It is indeed very doubtful if the microscopic method, entailing the proximity to the sol of a mass of metal, and consequent asymmetric flow of heat, can be successfully employed with small particles in vessels of macroscopic dimensions. We have frequently noticed, as would be expected from the smallness of the natural rate of fall of their particles, the extreme sensitiveness of distributed sols to slight changes of temperature.

There appears to be some uncertainty with regard to the factors which determine any departure from "ideal" behaviour which may actually occur, as they are stated sometimes in terms of "depth from the surface" and at others in terms of "concentration." We have observed, qualitatively, a gradation of colour in a fine gold sol at a depth of 2.5 cm. below the surface, and it seems reasonable to suppose that the determining factor is concentration—expressed as $1/(v - b)$ —rather than depth. The numerical concentration,

10^{11} per cubic centimetre, at which Costantin observed deviations, is not directly comparable with those studied with gold sols, since the actual volume concentration, approaching 1 per cent., was much greater in the former case.

In the experiments described in this paper the range of examination has been extended to a depth of 9 mm., and to a concentration of 10^{12} particles per cubic centimetre, and at these limits the behaviour cannot be distinguished from that predicted by Perrin's equation. Although the experiments are not sufficiently exact to enable small deviations, of the order of magnitude observed in the case of gases, to be detected, they show the complete absence of gross departures. The particles used have been those of ordinary colloidal solutions, of a size well below that corresponding with the supposed "critical radius." No theoretical significance can be attached to the latter magnitude, which means in practice the size of particles which are just prevented from falling by accidental movements in the liquid.

Summary.

Until 1920 the view that the numerical concentration of particles suspended in a liquid and in gravitational equilibrium, changes exponentially with depth was generally accepted. This view was based on the experiments of Perrin, Costantin, and Westgren, which showed the validity of the simple law over an extreme range of depth of 0.05 cm., and up to a numerical concentration of 5×10^{10} particles per cubic centimetre. Subsequent work by Burton, Porter and Hedges, and Barkas, has indicated that the simple law is not valid beyond a depth of 0.01 cm. from the surface, and that it may break down at a numerical concentration as low as 10^6 particles per cubic centimetre. The recent experiments are criticised on the grounds of inadequate temperature control and of failure to reach a state of equilibrium. A method of measuring the distribution of particles of colloidal gold over a range of 1 cm., while the suspension is maintained at a temperature varying by not more than 0.001° per hour, is described. Perrin's law has been found to hold good over a range of 0.9 cm. from the surface and up to a concentration of 10^{12} particles per cubic centimetre, at which limits no sensible departure from "ideal" behaviour was apparent.

Interferometric Measurements in the Spectrum of Krypton.

By C. V. JACKSON, Imperial College of Science and Technology.

(Communicated by A. Fowler, F.R.S.—Received May 27, 1932.)

Introduction.

The standard of length at present accepted by spectroscopists is the wave-length of the red line of cadmium, 6438·4696 Å., in dry air at N.T.P. This defines the angstrom, and although it probably differs from 10^{-10} metre by little more than one part in ten millions, it seems desirable to retain the present definition, since there is now some doubt as to the permanency of material standards of length. Moreover the present work shows that it is quite possible to compare wave-lengths with an accuracy of at least one part in fifty millions, or probably about ten times the accuracy which can be attained in the comparison of the lengths of a material standard and a light wave.

Unfortunately the red line of cadmium is not very convenient as a spectroscopic standard of length. It is situated in the deep red, is therefore somewhat difficult to photograph, and it is one of the fainter lines in the spectrum of cadmium. The Michelson lamp, which is the only strictly correct source of the radiation, requires to be heated in a furnace at a temperature of 320° C. This is highly undesirable since the heat from the furnace greatly increases the difficulty of making accurate interferometric wave-length comparisons. Finally the line is in a spectral region far removed from that where the majority of spectral lines occur, which is disadvantageous, since it is very much easier to make accurate comparisons of wave-lengths not far removed from each other.

In actual practice it is more convenient to use a group of lines of accurately determined wave-lengths, for determining the optical length of an étalon, than a single line. Hitherto, in some of the most accurate series of wave-length determinations the lines of neon have been used with a great measure of success as a substitute for the primary standard. It is noteworthy that although the wave-lengths of the neon scale are given to only seven figures, there is considerable evidence that, if a fair number of lines are used, this scale is probably correct to at least one part in thirty millions.

The spectrum of krypton contains a group of ten very strong lines in the violet which, on account of their position in the spectrum, appear to be even

more suitable than the neon lines. Further, since krypton is four times as heavy as neon, its lines are twice as sharp as those of neon. The lines are suitably spaced for the determination of the whole number of waves contained by an étalon, even when its approximate thickness is only very roughly known. The lines are easily produced by means of spectrum tubes which can be purchased commercially.

For these reasons the author has made an extensive series of interferometric observations on the wave-lengths of these lines. They appear to be very satisfactory, and it is hoped that the wave-lengths presented in this paper provide a scale which is accurate to about one part in fifty millions, or 0.0001 Å. The results show a mean systematic difference of this amount from some previous measurements of Humphreys, who, however, used the neon scale in place of the primary standard.

Experimental.

One of the interferometers used for this work was a new type of Fabry-Perot étalon designed by F. Twyman and constructed by Adam Hilger, Ltd. It is described as an "adjustable étalon," since it may be used with any plate separation between $\frac{1}{4}$ and 10 cm. It differs from most interferometers in the extremely high degree of accuracy to which the plates are worked, and also in the method of adjustment for parallelism of the plates, very fine pitched screws taking the place of the spring pressure devices previously used. The interferometer was used in conjunction with an E1 spectrograph, using the quartz train, since this gave adequate dispersion and at the same time permitted the red line of cadmium and the violet lines of krypton to be photographed with the same setting of the spectrograph.

The cadmium, krypton and neon lamps are carried on a small table which slides along a rail so that any one of them can be brought into position in front of a quartz condenser. This is of about 6 inches diameter and 15 inches focus and projects a five-fold magnified image on a screen, which contains a slit 1 inch high and $1/10$ th inch wide, accurately aligned with the spectrograph and the interferometer, so that an image centred on it is also centred on the slit of the spectrograph. It is placed at the focus of a quartz fluorite collimating lens, which parallelises the light before it passes through the interferometer, and ensures that strictly equivalent paths are taken by all the radiations passing through the étalon.

Except for the plates taken with very short paths (for determination of phase correction) a triple quartz fluorite achromat of 50 cm. focus was used

for projection of the rings. This lens is specially corrected to give a flat field, and appeared to cause no distortion of the ring system.

In order to ensure stability of the apparatus it is situated in the constant temperature room of the Imperial College. This is in the basement and has triple walls. It rests on a heavy independent concrete foundation, which protects the interferometer from vibration. The temperature changes in it are very slow and it is quite free from periodic day and night changes in temperature. The greatest change ever observed was about $1/10\text{th}^{\circ}\text{C}$. in the course of 3 hours. The humidity was fairly constant, varying between 40 per cent. and 60 per cent.

The first observations made with the adjustable interferometer were made on the spectrum of neon, with the object of determining its suitability for the accurate comparison of wave-lengths. It proved to be quite satisfactory for this purpose, when used in the constant temperature room, although it was not satisfactory when used in an ordinary laboratory. It was found that, if the étalon were allowed 3 or 4 hours to settle down after adjustment, the plate separation changed by only about two thousandths of a fringe per hour. The wave-length measurements of neon lines were in almost perfect agreement with the accepted secondary standards.

For the plates taken with the fixed separation étalons the apparatus was identical with that used previously by the author* for his work on the spectrum of iron. Separations of 1 cm. and 2 cm. were used, and the ring projecting objectives were of 13-inch and 6-inch focus, but the plates taken with the 6-inch lens were not suitable for measurement since the scale of the rings was too small. If the scale of the rings is too small it is almost impossible to set on the correct part of the ring segment (*i.e.*, the position of greatest density, which is not coincident with the centre). This leads to a systematic error in the wave-lengths, making them apparently greater than they are in reality. However, it was observed that the error was very regular, and even in an extreme case, taken specially to emphasise it by using a scale of projection that was much too small, it was less than $1/100\text{th}$ of a fringe, *i.e.*, a few units in the eighth figure of the wave-length. It therefore seems probable that no appreciable error from this source can possibly occur in the measurements made from the plates taken with large scale projection. If the standards were in the same part of the spectrum as the lines being measured, an error of this type would naturally have practically no effect on the accuracy of the resulting wave-lengths.

* 'Proc. Roy. Soc.,' A, vol. 130, p. 395 (1931), and vol. 133, p. 553 (1931).

It is of considerable importance to use ring objectives of about the correct focal length. For interference paths up to $\frac{1}{2}$ cm. a 6-inch lens should be used, for 1 or 2 cm. paths about 13-inch is best, while for paths above 3 cm. about 20-inch should be used.

The krypton tubes were prepared by Professor Lepape, and one of these supplied by Hilger sufficed for the whole of this work. It was found entirely satisfactory, and showed no sign of deterioration. The pressure of the krypton is about 6 mm. Hg., and the tube was excited with a small transformer, which was equally suitable for exciting the cadmium tube or a neon tube. The tubes could be used with equal convenience in the "end-on" position or in the ordinary position. The former was in general more satisfactory since the exposure required was very much less.

The cadmium lamp was of the type used by Michelson, and was used under the conditions recommended by him. It contained only a small quantity of cadmium to eliminate the risk of a pressure shift of the red line in the event of the tube becoming over-heated.

In order to reduce the exposure to the shortest possible time, Ilford Soft Gradation Panchromatic plates were used for the majority of the exposures.

The photographs were made by alternately exposing cadmium and krypton. Even with the shorter exposures eight changes were made, so that any unavoidable change in the length of the étalon during the course of an exposure could have only a very small effect on the wave-lengths obtained.

The Measurements.

The wave-lengths were obtained from measurements of 38 plates of the spectrum of krypton. Of these 20 were made with the adjustable étalon with a plate separation of 3 cm., 5 with a fixed étalon of 2 cm., and 13 with one of 1 cm. The plates taken with a separation of $\frac{1}{2}$ cm. were used only for the purpose of determining the phase correction. The measurement and reduction of the ring systems were carried out by the method used by the author (*loc. cit.*) in his work on the iron spectrum. No systematic difference between the values obtained from the first and the second rings was noticed in any of the three series. On account of the variation in intensity of the lines it was not possible to measure all of them on every plate. The observations were corrected to N.T.P. but it was not possible to make any correction for the humidity of the air. This had a mean value of 50 per cent., varying between about 40 per cent. and 60 per cent. of saturation.

The mean probable error of the wave-lengths determined with 3 cm. separation was found, from the agreement of the results obtained from different plates of this series, to be about ± 0.0001 A. For the determinations with the 2 cm. and with the 1 cm. étalon it was found to be about ± 0.0002 A. The fact that the results obtained from the plates taken with a 1 cm. separation are as accurate as those taken with a separation of 2 cm. is explained by the far greater number of observations made in the former case.

The agreement between the results of the three series of measurements is very good, the differences being no greater than the probable errors of the measurements, as regards both systematic and accidental errors. This may be regarded as a definite indication that the apparent wave-lengths of the lines are independent of the separation of the interferometer plates. This is a matter of great importance in selecting a group of spectral lines for use as standards of the highest accuracy.

The results of the measurements are given in Table I together with those of

Table I.

Jackson, mean.	No. of Plates and P.E.	Jackson.			Humphreys.	Meggers.
		3 cm.	2 cm.	1 cm.		
4273.9702	26 A	9702	9704	9700	9705	9696
4282.9689	14 A	9689	9687	9690	9686	967
4318.5522	7 A	5522	—	—	5523	552
4319.5801	21 B	5801	580*	580*	5798	580
4362.6425	24 A	6425	6425	6425	6429	6422
4376.1221	28 A	1221	1224	1220	1217	122
4399.9673	16 A	9673	9671	9675	9675	969
4453.9179	28 A	9179	9182	9176	9183	9174
4463.6906	29 A	6905	6906	6907	6897	690
4502.3548	28 B	3546	3551	3548	3546	354

* These measurements were disturbed by λ 4318.

Humphreys† and of Meggers,‡ which are given in the last two columns. The first column gives the mean of the three values obtained by the author. In evaluating these means, weights of three, two and one were assigned respectively to the series obtained with separations of 3, 2 and 1 cm. The second column gives the number of plates on which the line was measured, and also the letter A or B. A signifies that the probable error of the wave-length is less than 0.0001 A., and B that it is less than 0.0002 A.

† 'J. Res. Bur. Stand. Wash.,' vol. 5, p. 1041 (1930).

‡ 'Sci. Pap. Bur. Stand. Wash.,' vol. 17, p. 193 (1921).

There is good agreement between the present wave-lengths and those of Humphreys, the mean accidental difference being only 0.0003 A., while the mean systematic difference ($J - H$) is only 0.0001 A. The measurements of Humphreys were referred to the neon scale, and therefore the excellent agreement with the results of the present investigation, a direct comparison with the primary standard, indicates that the neon scale is practically equivalent to the primary standard, even when one is working to eight figures. The preliminary work on the neon scale, which was carried out by the author in order to test the new interferometer, also indicates that wave-lengths measured with the neon scale are equivalent to direct comparisons with cadmium, to within one or two ten-thousandths of an angstrom.

The agreement with the results of Meggers, though not as good as that with Humphreys, may be considered quite satisfactory considering the fact that most of Megger's wave-lengths are given to only seven figures and, moreover, were derived from the measurement of only a small number of plates. The mean accidental difference is ± 0.0007 A., while the mean systematic difference ($J - M$) is $+0.0003$ A.

The lines were also observed with a plate separation of 6 cm. and still found to give quite good fringes, but the plates were not measured. The interferometer plates were sputtered with platinum, which is very suitable for wave-length determinations on account of the constancy of the phase correction.

The Yellow and Green Lines of Krypton.

In addition to the group of lines in the violet, the spectrum contains three lines in the yellow and green which are the most prominent lines in the spectrum, of wave-lengths 5870, 5570 and 5562 A. The wave-lengths of these lines were determined, with the same care as those in the violet. The results, however, as can be seen from Table II, are not in good agreement with those of other observers, the differences being very much greater than would be expected

Table II.

Jackson, 1932.		Pérard.		Humphreys, 1930.	Meggers, 1921.	Fabry and Buisson, 1913.
		1932.	1923.			
5562.2266	25 B	22576	2257	2251	224	—
5570.2899	31 B	2894	2892	2890	2872	2908
5870.9167	32 B	9161	9154	9153	9137	9172

from the probable errors of the various observations. For this reason the lines appear to be unsuitable for use as standards of the highest accuracy. This is somewhat surprising in view of P  rard's* observations on the line 5562. He gives the wave-length of the line to nine figures, and assesses the probable error at ± 0.000025 A. ; he considers that this line should be adopted as a standard of length, on account of the very high degree of accuracy with which he believes its wave-length can be measured.

In conclusion the author wishes to express his gratitude to Professor A. Fowler for his continual interest, and for the very valuable advice and encouragement which he gave so generously throughout the course of the work.

Summary.

The wave-lengths of the ten bright violet lines of the first spectrum of krypton have been measured, with an accuracy of about ± 0.0001 A., by interferometric comparison with the primary standard. The systematic error is probably considerably less than 0.0001 A. The agreement with previous measurements by Humphreys on the neon scale is very good, the mean systematic difference ($J - H$) being $+0.0001$ A., while the mean accidental difference is ± 0.0003 A.

This group of lines appears to be very suitable as a standard of length for spectroscopic purposes. It is situated in a very convenient part of the spectrum, and the lines have a very high degree of homogeneity, about twice that of the neon lines. With a suitable tube the lines can be photographed with a short exposure. They are suitably spaced for the evaluation of the thickness of an   talon, even when its approximate value is in error by several hundred wave-lengths.

In addition to the violet lines the three brightest lines, $\lambda\lambda$ 5870, 5570 and 5562, of krypton were measured, but they are not considered suitable as standards of the highest accuracy because the wave-lengths found by the various observers differ by amounts very much greater than their respective probable errors.

The neon scale has been carefully tested, and it appears that if a fair number of standards are used this scale gives results identical with those obtained by the use of the primary standard to within one or two units in the eighth figure.

* 'C. R. Acad. Sci. Paris,' vol. 194, p. 1633 (1932).

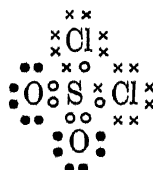
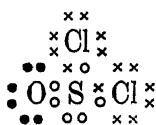
*Dipole Moments and Molecular Structure. Part III.—The
Oxychlorides of Sulphur.*

By JOHN WILLIAM SMITH, D.Sc., Ph.D., The Sir William Ramsay Laboratories
of Inorganic and Physical Chemistry, University College, London.

(Communicated by Professor F. G. Donnan, F.R.S.—Received May 28, 1932.)

Introduction.

The work described here forms part of a study of the dipole moments of simple inorganic compounds. Polarisation measurements upon thionyl chloride and sulphuryl chloride appeared to be of particular interest in view of the fact that in these compounds the oxygen-sulphur linkages must be in the nature of semi-polar double bonds, since both members of the electron pairs concerned in these linkages come from the sulphur atoms, thus :



where ●, ○ and × represent electrons derived originally from the oxygen, sulphur, and chlorine atoms respectively. Consequently, it would be anticipated that these compounds might show relatively large dipole moments, and would also possess a considerable atomic polarisation. (A communication concerning the relationship between atomic polarisation and molecular structure is being published shortly.) No previous measurements of the polarisations of these compounds, either in the gaseous state or in dilute solution, appear to have been recorded.

Materials.

The thionyl chloride and sulphuryl chloride were prepared by fractionating the "pure" commercial products several times. The boiling points of the materials thus produced were 77° C. and 70° C. respectively. The benzene was purified as described in Part II of this series,* and was dried over successive quantities of sodium wire until the latter retained its metallic lustre.

* 'Proc. Roy. Soc.,' A, vol. 136, p. 256 (1932).

The Measurements.

Series of dilute solutions of the thionyl chloride in benzene and sulphuryl chloride in benzene were prepared. The dielectric constants of these solutions were measured at 25° C. and 45° C. for waves of frequency $1,227,282 \pm 20$ cycles per second, by the method described in Part I,* Hartshorne and Oliver's value for the dielectric constant of benzene† being taken as standard value. Using the purified chlorides and benzene which had been well dried as described above, no serious trouble was encountered through any reaction between the solute and the silvered condenser plates, although less pure materials attacked them rapidly.

In order to compare the molecular polarisations determined at Hertzian wave-lengths with the molecular refractions of the liquids, the refractive indices of the solutions were measured. A Pulfrich refractometer was used for this purpose, and for convenience the determinations were carried out at 15° C. and referred to the mercury green line (5461 Å.). The densities of the solutions were also measured at 15°, 25°, and 45° C., and the molecular polarisations and molecular refractions of the solutes were calculated from this data according to the modified form of the Clausius Clapeyron and Lorentz-Lorenz equations respectively, which are applicable to dilute solutions.

Results with Thionyl Chloride.

The experimental data obtained with the solutions of thionyl chloride are summarised in Table I, in which f represents the molecular fraction, M the molecular weight, ϵ the dielectric constant, ρ the density, P the molecular polarisation, n the refractive index, and R the molecular refraction, whilst the suffixes 1 and 2 refer to the solute and solvent respectively.

The difference between the values of P_1 at infinite dilution at 25° C. and at 45° C. respectively represents the change in the value of the B/T term in the simplified form of the Debye equation $P = A + B/T$, where $B = 4\pi N\mu^2/9k$ (N is the Avogadro number, μ the dipole moment of the molecule, and k is Boltzmann's gas constant. Since $\delta P = 2.5$, we find that $B = 2.5 \times 298 \times 318/20 = 11,850$, and $\mu = 0.01272 \sqrt{B} \times 10^{-18} = 1.38 \times 10^{-18}$ e.s.u. From this data it follows that at 25° the orientation polarisation (B/T) is 39.8 c.c., and A is 34.2 c.c. This value of A should represent the sum of the electronic and atomic polarisations, P_E and P_A . The former of these should

* 'Proc. Roy. Soc.' A, vol. 136, p. 251 (1931).

† 'Proc. Roy. Soc.' A, vol. 123, p. 664 (1929).

Table I.

	Benzene.	I.	II.	III.	IV.
f_1	—	0.02000	0.03996	0.07852	0.12934
$M_1f_1 + M_2f_2$	78.000	78.820	79.638	81.219	83.303

Temperature 25° C.

ϵ	2.2725	2.346	2.416	2.553	2.732
ρ	0.8738	0.8863	0.8990	0.9233	0.9548
$P_1f_1 + P_2f_2$ (c.c.)	26.59	27.53	28.41	30.01	31.93
P_2f_2 (c.c.)	26.59	26.05	25.52	24.50	23.15
P_1f_1 (c.c.)	—	1.48	2.89	5.51	8.78
P_1 (c.c.)	—	73.8	72.2	70.2	67.9

Extrapolated value of P_1 for $f \rightarrow 0$ is 74.0 c.c.

Temperature 45° C.

ϵ	2.233	2.297	2.361	2.486	2.649
ρ	0.8521	0.8638	0.8755	0.8995	0.9309
$P_1f_1 + P_2f_2$ (c.c.)	26.664	27.54	28.39	29.92	31.74
P_2f_2 (c.c.)	26.664	26.13	25.60	24.57	23.21
P_1f_1 (c.c.)	—	1.41	2.79	5.35	8.53
P_1 (c.c.)	—	70.6	69.8	68.1	65.9

Extrapolated value of P_1 for $f \rightarrow 0$ is 71.5 c.c.

Temperature 15° C.

n	1.50936	1.50978	1.51020	1.51077	1.51140
n^2	2.2782	2.2794	2.2807	2.2824	2.2843
ρ	0.8842	0.8969	0.9098	0.9345	0.9663
$R_1f_1 + R_2f_2$ (c.c.)	26.355	26.274	26.185	26.027	25.842
R_2f_2 (c.c.)	26.355	25.828	25.302	24.286	22.947
R_1f_1 (c.c.)	—	0.446	0.883	1.741	2.895
R_1 (c.c.)	—	22.3	22.1	22.17	22.30

Mean value of $R_1 = 22.2$ c.c.

be approximately identifiable with the molecular refraction R_1 , extrapolated for infinite wave-length. The value 22.2 c.c. which is obtained as a mean value for the molecular refraction of thionyl chloride in solution is well in agreement with the value 22.34 c.c. for the pure liquid, calculated for light of the same wave-length (5461 Å.) by interpolation of the refractive index data of Nasini,* and using the density formula of Sugden, Reed and Wilkins.† Consequently it appears reasonable to conclude that the value of R_1 for infinite

* 'Rend. Acc. Linc.,' vol. I, p. 76 (1885).

† 'J. Chem. Soc.,' vol. 127, p. 1538 (1925).

wave-length will be equal to the molecular refractivity calculated for the pure liquid, viz., 21.0 c.c. There is obviously a great discrepancy between the values of A and R, and the difference, 13.2 c.c., must be attributed to atomic polarisation. Although the absolute value of the latter as deduced in this manner may be subject to some error, the evidence certainly indicates that the atomic polarisation is by no means negligible in the case of thionyl chloride.

If, on the other hand, the atomic polarisation be neglected, and the orientation polarisation be calculated in the manner usually adopted as equal to the difference between the molecular polarisation and the molecular refraction for the mercury green line, we find that at 25° C. $B = 15,436$, and $\mu = 1.58 \times 10^{-18}$ e.s.u. Thus in cases of this type considerable errors can be introduced into the dipole moment calculation by neglecting the atomic polarisation term.

Results with Sulphuryl Chloride.

The significant data for the solutions of sulphuryl chloride in benzene are collected in Table II. The lettering has exactly the same significance as in Table I.

Table II.

	Benzene.	I.	II.	III.	IV.
f_1	—	0.02221	0.04422	0.07376	0.09650
$M_1f_1 + M_2f_2$	78	79.266	80.521	82.261	83.501

Temperature 25° C.

ϵ	2.2725	2.379	2.488	2.635	2.753
ρ	0.8738	0.8893	0.9050	0.9260	0.9417
$P_1f_1 + P_2f_2$ (c.c.)	26.59	28.07	29.50	31.34	32.70
P_2f_2 (c.c.)	26.59	26.00	25.41	24.63	24.02
P_1f_1 (c.c.)	—	2.07	4.09	6.71	8.68
P_1 (c.c.)	—	93.3	92.5	91.0	90.0

Extrapolated value of P_1 for $f \rightarrow 0$ is 94.3 c.c.

Temperature 45° C.

ϵ	2.233	2.329	2.425	2.560	2.658
ρ	0.8521	0.8668	0.8814	0.9020	0.9169
$P_1f_1 + P_2f_2$ (c.c.)	26.664	28.07	29.42	31.20	32.42
P_2f_2 (c.c.)	26.664	26.07	25.48	24.70	24.09
P_1f_1 (c.c.)	—	2.00	3.94	6.50	8.33
P_1 (c.c.)	—	90.1	89.1	88.1	86.3

Extrapolated value of P_1 for $f_1 \rightarrow 0$ is 90.8 c.c.

Table II—(continued).

	Benzene.	I.	II.	III.	IV.
Temperature 15° C.					
n	1.50936	1.50834	1.50730	1.50575	1.50476
n^2	2.2782	2.2751	2.2720	2.2673	2.2643
ρ	0.8842	0.9001	0.9161	0.9375	0.9536
$R_1 f_1 + R_2 f_2$ (c.c.)	26.355	26.266	26.171	26.058	25.962
$R_2 f_2$ (c.c.)	26.355	25.770	25.190	24.411	23.812
$R_1 f_1$ (c.c.)	—	0.496	0.981	1.647	2.150
R_1 (c.c.)	—	22.24	22.18	22.31	22.28

Mean value of R_1 is 22.25 c.c.

Calculating the orientation polarisation from the change of P_1 with change of temperature, in the same manner as was adopted in the case of thionyl chloride, we find that δP (25°–45° C.) = 3.5 c.c., $B = 3.5 \times 298 \times 318/20 = 16,584$, and $\mu = 1.64 \times 10^{-18}$ e.s.u.

The figure 22.25 c.c. obtained for the mean value of the molecular refraction of sulphuryl chloride in benzene solution for the mercury green line, 5461 Å., is rather higher than the value (21.41 c.c.) calculated for this wave-length from the refractive index and density of the pure liquid as observed by Nassini and Costa.* The difference is so small, however, that it is permissible to take the value of the molecular refractivity for infinitely long waves as equal to that calculated for the pure liquid by extrapolation of Nasini and Costa's measurements by means of Cauchy's formula, viz., 20.74 c.c.

The value of the orientation polarisation at 25° C. is $16,584/298 = 55.6$ c.c. Since the total molecular polarisation at this temperature is 94.3 c.c., $P_E + P_A = 38.8$ c.c. The difference between this figure and the molecular refraction for infinitely long waves must be attributed to atomic polarisation. The latter is therefore given by $38.8 - 20.74 = 18.1$ c.c.

If, on the other hand, the atomic polarisation be neglected, and the orientation polarisation be taken as equal to the difference between the molecular polarisation and the molecular refraction for the mercury green line, the value obtained for B is $298 \times 72 = 21,456$, and $\mu = 1.86 \times 10^{-18}$ e.s.u.

Discussion of Results.

The fact that the semi-polar sulphur-oxygen linkage tends to lead to rather high values of the dipole moment has already been observed by Bergmann,

* 'Pubb. Inst. Chim.,' Rome, p. 111 (1891).

Engel and Sandor* and by de Vries and Rodebush† for the series diphenyl sulphide, diphenyl sulphoxide, and diphenyl sulphone. A comparison of their results with those found in the present investigation is very interesting. Owing to the fact that sulphur dichloride dissociates so very readily‡, it is impossible in the chloride series to find with any certainty the effect of the addition of the oxygen atom to the SCl_2 group. A rough estimate, however, may be made of the electric moment of the sulphur dichloride molecule by taking the value of 27 c.c. deduced by Lowry and Jessop for the molecular polarisation of sulphur dichloride in solution in chlorine.§ Calculating the molecular refraction from their value of the density at 15°C . (*circa* 1.629) and the refractive index (of the presumably impure compound) for the sodium D line (1.557), we obtain a value of 20.3 c.c. Whence $P_E + P_A = 6.7$ c.c., and $\mu = 0.56 \times 10^{-18}$ e.s.u.

A comparison of the electric moments in the two series shows the following :

	$\mu \times 10^{18}$		$\mu \times 10^{18}$
Sulphur dichloride		Diphenyl sulphide	
Cl_2S	[0.56]	Ph_2S	1.47 1.565¶
Thionyl chloride		Diphenyl sulphoxide	
Cl_2SO	1.38	Ph_2SO	4.08 4.17¶
Sulphuryl chloride		Diphenyl sulphone	
Cl_2SO_2	1.64	Ph_2SO_2	— 5.05¶

The big differences between the moments of thionyl chloride and diphenyl sulphoxide, and between those of sulphuryl chloride and diphenyl sulphone is very striking. However, it must be remembered that in the case of the chlorides there are two strongly negative atoms towards the side of the sulphur atom remote from the oxygen atom, which tend to neutralise the moment produced by the semi-polar sulphur-oxygen linkage, in which a very strong dipole would be expected to be located, owing to the oxygen atom sharing an electron-pair which comes from the sulphur atom. If the three atoms adjoining the sulphur atom, the two chlorine atoms and the one oxygen atom, in thionyl chloride be assumed to be situated at the corners of the base of a

* 'Z. phys. Chem.,' B, vol. 10, p. 397 (1930).

† 'J. Amer. Chem. Soc.,' vol. 53, p. 2888 (1931).

‡ Lowry and Jessop, 'J. Chem. Soc.,' p. 1421 (1929).

§ 'J. Chem. Soc.,' p. 782 (1930).

|| Bergman, Engel and Sandor (*loc. cit.*).

¶ de Vries and Rodebush (*loc. cit.*).

(probably irregular) tetrahedron, with the sulphur atom at the apex, the observed electric moment will be the resultant of three moments along the edges joining the apex to the corners of the base of the tetrahedron, and of these moments that along the line of the sulphur-oxygen linkage will be the greatest. The latter moment will act at a wide angle to the resultant of the other two, and, in consequence, the latter may reduce its magnitude very considerably.

In the diphenyl compounds, however, the electric doublet in the Ph_2S grouping is probably in the same sense as that in the sulphur-oxygen linkages, with its negative end towards the sulphur atom. Consequently in this case the two components of the moment tend to aid one another. Again, in this case the large positive charge produced on the sulphur atom in the presence of a semi-polar oxygen-sulphur linkage will cause a considerable polarisation of the readily polarisable phenyl groups adjacent to it, thus tending to increase the moment of the molecule as a whole. These effects will be magnified still further on the addition of the second semi-polar oxygen atom to form diphenyl sulphone. This would naturally make the difference between the moments of diphenyl sulphone and diphenyl sulphoxide much greater than that between the moments of thionyl chloride and sulphuryl chloride. The latter difference is surprisingly small, however, and the evidence would seem to suggest that not only is the Cl-S-Cl angle in the chlorides much narrower than the Ph-S-Ph angle in the diphenyl compound, but also that the O-S-O angle may be wider in the former than in the latter.

In view of the fact that an indication has been obtained that the oxychlorides of sulphur exhibit rather a large atomic polarisation (about 13.2 c.c. for thionyl chloride and 18.1 c.c. for sulphuryl chloride), it is interesting to note that a very high value (33.4 c.c.) has also been observed with dimethyl sulphate* which also contains semi-polar bindings, and a fairly high value (8 c.c.) with nitrobenzene. This tends to confirm the view that a high atomic polarisation is a common property of the semi-polar linkage. If this be agreed, the fact that sulphur dioxide has at most only quite a low atomic polarisation (0.9 c.c.)†, appears to support the view that the sulphur dioxide molecule contains no true semi-polar linkage.

The only alternative explanation of the great discrepancies between the values of A (in the equation $P = A + B/T$) and of R_1 is that the dipole moments increase somewhat with rise in temperature. The only way to distinguish

* Smythe, 'J. Amer. Chem. Soc.', vol. 51, p. 2051 (1929).

† Errera, "Polarisation Diélectrique," Paris (1928).

between these two hypotheses would be to determine the molecular polarisation of thionyl chloride and of sulphuryl chloride in the solid state. At present, however, no data bearing upon this subject appear to be available.

In conclusion, the author desires to express his indebtedness to Professor F. G. Donnan, F.R.S., for the continued interest which he has taken in these investigations.

Summary.

The dielectric constants and densities of benzene solutions of thionyl chloride and sulphuryl chloride have been measured at 25° and 45° C., and the refractive indices and densities of the same solutions determined at 15° C. From these data the corresponding molecular polarisations and molecular refractions of the two chlorides have been calculated. The electric moments calculated from the change in the molecular polarisation with change in temperature are 1.38×10^{-18} and 1.64×10^{-18} respectively; and those calculated from the molecular polarisation at 25° and the molecular refraction are 1.58×10^{-18} and 1.86×10^{-18} respectively. The discrepancy between these two series of values indicates either that the atomic polarisations of these molecules is high (about 13.2 c.c. for thionyl chloride and 18.1 c.c. for sulphuryl chloride) or that the dipole moment of these compounds increases with rise of temperature. The results are discussed with reference to the related series diphenyl sulphide, diphenyl sulphoxide, and diphenyl sulphone.

The Swelling of Charcoal. Part II.—Some Factors Controlling the Expansion Caused by Water, Benzene and Pyridine Vapours.

By D. H. BANGHAM, N. FAKHOURY, and A. F. MOHAMED, The Egyptian University, Cairo.

(Communicated by D. L. Chapman, F.R.S.—Received June 7, 1932.)

1. *Introductory.*

The experiments described in Part I of this series* led to the conclusion that the expansion which charcoal undergoes in contact with water vapour, carbon dioxide, ammonia, and sulphur dioxide is connected with the existence in the adsorbed phase of a pressure due to the free movement of the molecules in directions parallel to the interface. In a later paper† it was pointed out that if the relation between the percentage linear expansion (x) and the surface pressure (F) be assumed to take the form

$$x = \lambda F, \quad (1)$$

where λ is a constant, then the equation found empirically to relate the adsorption and expansion in the case of these gases reduces to an expression of the Schofield-Rideal equation of state :

$$F(A - B) = iRT. \quad (2)$$

Here A stands for the area per (gram) molecule, and is equal to $M\Sigma/s$ if s is the weight in grams of an adsorbate (of molecular weight M) taken up by 1 gram of charcoal the effective surface area of which is Σ ; B is the "incompressible" cross-sectional area per gram molecule; and i is a constant determined by the mutual cohesion of molecules and is equal to unity when such cohesion is absent.

The experimental isotherms of water, carbon dioxide, and sulphur dioxide in contact with the charcoal then used, later referred to as charcoal I, gave tolerably good agreement with a theoretical equation derived by applying the Gibbs relation to a surface phase obeying equation (2).‡ If it is assumed that the lowering of the surface energy attending the adsorption is completely

* 'Proc. Roy. Soc.,' A, vol. 130, p. 81 (1930).

† 'J. Chem. Soc.,' p. 1324 (1931).

‡ Agreement was not obtained in the case of ammonia.

determined by the magnitude of the surface pressure, the Gibbs equation becomes, when expressed in terms of symbols used,*

$$d \log_e p = \frac{M\Sigma}{RT} \cdot \frac{1}{s} \cdot dF. \quad (3)$$

If equation (1) is also valid this becomes

$$d \log_e p = \frac{M\Sigma}{\lambda RT} \cdot \frac{1}{s} \cdot dx, \quad (4)$$

an equation in which all the variables are experimentally determinable. Equation (4) receives some support from the data for benzene and pyridine in contact with charcoal II, a charcoal used in the present investigation.† The data for water, on the other hand, do not give any extended range of agreement with this equation. A reason for this becomes apparent on examining the characteristic curves given in fig. 5A, where values of Mx/s , the percentage expansion per gram molecule per gram of charcoal, later referred to as the "molecular expansion," are plotted against the expansion itself. If equation (1) is valid the graphs should be comparable with the $FA - F$ graphs commonly used in surface chemistry. Judged from this standpoint the experimental results for water give a clear indication that condensation of the film from a gaseous to a liquid state is taking place: for not only does x remain constant over a considerable range of s -values, but the rising parts of curves 2, 3 and 4, if produced back to meet the ordinate axis, would do so very close indeed to the origin. Rideal‡ has emphasised that equation (3) is invalid when two surface phases co-exist.

With regard to films in the gaseous condition, it is important to note that unless future experiment should prove λ and Σ to be independent of the size and shape of the molecules, the molecular expansion at zero adsorption—unlike the product F_0A_0 —must vary from substance to substance.§ Since $s_0/M = \Sigma/A_0$ and $F_0A_0 = RT$ its value is given by

$$Mx_0/s_0 = M\lambda F_0/s_0 = \lambda RT/\Sigma. \quad (5)$$

Experimentally, Mx_0/s_0 could be determined either (i) by extrapolation based

* *Loc. cit.*, p. 1327.

† It is similarly supported by data (to be published later) for methyl, ethyl, *n*-propyl, isopropyl, and *n*-butyl alcohols.

‡ "Introduction to Surface Chemistry," Cambridge, p. 52 (1930).

§ Bangham and Fakhoury, *loc. cit.*, p. 1326.

on measurements of very small expansions, or (ii) by the use of equation (4) to determine λ/Σ . Combining (4) and (5) it is seen that

$$\frac{Mx_0}{s_0} = \frac{Mdx}{s d \log_e p} = 0.4343 \frac{Mdx}{s d \log_{10} p}. \quad (6)$$

Equation (6) permits the calculation of Mx_0/s_0 from data near to saturation. An excellent test of the accuracy of (1) would therefore be afforded by a comparison of the results obtained by these two methods.* Unfortunately, not only does method (ii) fail in the case of water, but method (i), for reasons which will be discussed, appears equally valueless in the case of benzene and pyridine. Whether equation (1) expresses an exact or an approximate relation it is therefore not within the scope of the present investigation to decide; but because our results lend themselves most readily to interpretation in terms of this equation, the practice of recording them by means of the characteristic molecular expansion graphs will be adhered to.

2. *Experimental Materials and Method.*

The difficulties experienced in standardising charcoal and obtaining reproducible results therefrom are much enhanced when the charcoal has to be used in the form of a block. Many of the recognised methods of treatment lead to disintegration of the blocks; moreover, unless the carbonisation has been carried out with superlative care, even specimens cut from the same block cannot be expected to behave alike. The charcoals used in our experiments were cut from a block of pinewood charcoal kindly supplied by Dr. F. Meehan; this had been carbonised under carefully controlled conditions which have been fully described.† Since water was to be one of the substances to be used as an adsorbate, it was considered advisable to reduce the ash-content as much as possible. The charcoal, cut into sections, was steeped in hydrofluoric acid for 15 days and afterwards washed with conductivity water for 47 days.

Rough measurements of the expansion produced on immersing, in different liquids, pieces of the charcoal which had been variously treated brought to light the important fact that vacuum-heating above a certain temperature brings about a considerable reduction in this effect. To account for this, and for the simplicity of the water, carbon dioxide, and sulphur dioxide iso-

* The test is not actually a crucial one, on account of the uncertainty attaching to the application of the Gibbs equation to such a surface as that of charcoal.

† Meehan, 'Proc. Roy. Soc.,' A, vol. 115, p. 199 (1927).

therms of charcoal I,* we made the tentative assumption that the activation of natural wood charcoal by vacuum treatment at high temperatures leads to the formation of condensed films† when adsorption takes place. In preparing the charcoal for the present series of investigations we aimed at giving it such treatment as would render it as free as possible from gross impurity and from loosely held volatile matter, but would be unlikely to give it the properties of an active charcoal. It was found that vacuum-heating to 560° C., while sufficing to remove in time the tarry matter of which the charcoal originally contained a large proportion, had little effect on the immersion values of the expansion. After being dried in an air oven at low temperature, therefore, Meehan's charcoal was subjected to 14 hours' pumping at 560° C. Thus prepared, the charcoal had a bulk density of 0.18 gram per cubic centimetre and an ash-content of 0.33 per cent. The ash was brown, acid soluble, and contained much iron.

Charcoal II, the piece used in all experiments except those of the last two series with water vapour, was a square prism 49 mm. long, weighing 0.498 gm.

Charcoal IIa, used in the water vapour experiments of Series II and III, was from the same original stock; its treatment is described in a later section.

The Benzene.—Kahlbaum's thiophene-free benzene was thrice solidified and dried by long standing over finely divided sodium. On redistillation the boiling point was constant at 80.2° C. After vacuum sublimation in the apparatus the melting pressure was 36.10 mm.

The Pyridine.—Kahlbaum's purest pyridine was allowed to stand for a long period over freshly ignited barium oxide. It was then thrice solidified, and distilled over freshly ignited barium oxide. When redistilled its boiling point, constant to 0.1°, was 115.0° C. at 759 mm.‡ The product was repeatedly resublimed after its introduction into the apparatus.

The Extensometer.—The extensometer used in most of the experiments was a nickel-coated brass instrument of the same general design as that described in Part I; the magnification (*i.e.*, the ratio of the movement of the pointer to that of the charcoal) was, however, rather lower, *viz.*, 33.1. Such instruments have been found to give reliable results provided the knife-edges are reasonably strong and far part; this condition sets a limit to the magnification that can

* Part I.

† *I.e.*, films other than gaseous; it appears probable that the high temperature treatment of the charcoal would favour the "secondary" adsorption process (see sections dealing with benzene and pyridine) which causes the charcoal to contract again.

‡ *Cf.* van der Meulen and Mann, 'J. Amer. Chem. Soc.', vol. 53, p. 451 (1931), according to whose interpolation formula the boiling point at 759 mm. is 115.04° C.

be obtained with a pointer of convenient length. Observations were made with a vernier microscope reading to 0.01 mm. against a carefully engraved and previously calibrated scale attached to the extensometer; no correction was necessary, therefore, for refraction through the curved wall of the containing tube. The length of the charcoal being 49 mm., it follows that a reading error of 0.01 mm. would have led to one of $100 \times 49^{-1} \times 33.1^{-1} \times 0.01 = 0.0006$ in the corresponding value of x , the percentage linear expansion. The fact that when once the conditions for obtaining reproducible results had been determined, the early observations (*i.e.*, those corresponding to small adsorptions) were found to agree within this order of accuracy provides a sufficient assurance that the extensometer was free from the zero-point error referred to in Part I. On account of the imperfection of the knife-edges and the slight slipping of the plate between them which must necessarily take place when their distance apart is increased by large expansions, the error of measurement increases with the expansion. From the behaviour of extensometers similarly constructed, but provided with a special calibrating device, this error is estimated to be of the order of 1 per cent. of the total movement in the case of the largest expansions met with in the experiments described. The error is systematic, the measured x -values being too low when x is great. No correction was applied.

The Apparatus.—Fig. 1 is a sketch of the apparatus in its final form. To avoid the use of taps, its constituent parts were separated by mercury cut-outs. The bulb B contained the stock of the substance under investigation and the two adjacent U-tubes were provided for its vacuum sublimation, to free it from dissolved gases; during this process one or other of the U-tubes was immersed in solid carbon dioxide, and the cut-out A was open to the pumps.

The measuring system consisted of the triple gas burette C_1, C_2, C_3, C_4 , and the gauge F, F_1, F_2 , which also served as a double cut-out. The two limbs of the gauge were of the same tubing, carefully chosen and tested for uniformity. The gauge readings were made with a reliable cathetometer reading to 0.01 mm. Corrections for difference of meniscus height in the two limbs of the gauge were applied throughout. The vacuum limb of the gauge was connected with a McLeod gauge, and through the tap G with the train of pumps. The tap was provided in order that constructional alterations—such as replacing the pentoxide tube—could be effected in the pumping system without admitting air to the whole apparatus; since it was exposed to the experimental vapour only when the pumps were running, there was no obvious objection to its use in this position.

After measurement in the burette, each quantity of vapour was admitted through the cut-out F_1 to the extensometer vessel H. To enable this to be done without danger of breaking the mercury column at the junction, the appendix D was momentarily cooled with a freezing mixture to equalise the pressures while the mercury was being lowered. The appendix was also useful in carrying out desorption experiments.

Calibration.—At the calibration stage the appendix D was provided with a tap (later sealed off) through which hydrogen could be admitted from its purifying train. During the calibration (and all subsequent measurements of

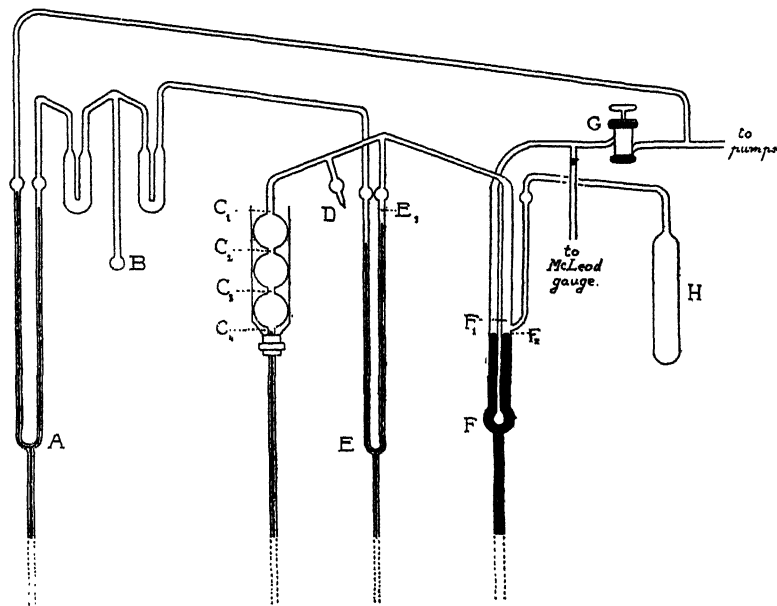


FIG. 1.

new quantities of vapour in the burette) the mercury in the gauge was set to a mark at F_1 and that in the narrow cut-out E E_1 to a mark at E_1 . Calibration of the burette system was then effected by setting the mercury in turns to each of the etched marks C_1 , C_2 , C_3 , C_4 , and reading the corresponding pressures. The same procedure was carried out in the case of the combined system; burette *plus* extensometer vessel; during this calibration and all subsequent measurements of vapour in the combined system the mercury in the right-hand limb of the burette was set to the mark F_2 .

Procedure.—Using the mercury cut-outs in the manner indicated it was possible to introduce to the charcoal successive quantities of vapour, each amounting to some two or three normal cubic centimetres. As long as the

pressure of unadsorbed vapour in the extensometer vessel remained fairly low, the mercury in the gauge was raised well above F_1 as soon as the necessary readings had been taken, and a fresh quantity of vapour admitted to the burette from the storage bulb. When, however, the pressure rose to values near saturation the procedure was varied: the mercury was raised in turn to each of the marks C_3 , C_2 , C_1 , and complete series of readings taken in each case. Connection between the burette and extensometer vessel was broken only when, the mercury in the former standing at C_1 , further increase of pressure could not be produced except by the admission of more vapour from the storage bulb. This procedure reduced the number of transfer operations necessary to reach saturation and made for better control of the approach to this condition.

Control of Temperature Variations and other Sources of Error.—To minimise the errors due to fluctuations of room temperature different means were used in different cases. The proximity of the reading instrument made it impossible to bath the extensometer tube in anything larger than a Dewar vessel, and, in cases where a fluctuation of a degree or so had an insignificant effect on the results, uniformity of temperature throughout the system was aimed at, rather than the absolute constancy of the temperature of its major members. The burette was enclosed in a small water-jacket which would rapidly take up the temperature of the surrounding air. By suitably adjusting the screens which surrounded the apparatus it was generally possible to keep the temperature of the whole system, including the connecting tubing, constant to within two or three tenths of a degree and—although some of the experiments were rather protracted—to complete each run with no greater departure than 1.5°C . from the mean day-to-day working temperature.

In the second and third series of experiments with water, in view of the smallness of the expansions recorded, a closer control of the temperature was needed in order to eliminate errors arising from the differential expansion of the charcoal and the extensometer. The temperature of the extensometer vessel was here maintained constant throughout each run to a small fraction of a degree. In these, and in the experiments with benzene where the charcoal was heated between each set of readings, great care was exercised to ensure that the extensometer had taken up the temperature of its surroundings before any readings were made. In the absence of vapour the extensometer reading formed a useful means for determining the temperature of the charcoal and estimating the time-lag.

Tests of the validity of Boyle's law, carried out in the burette system, showed this law to be applicable without serious error to the vapours of benzene

and pyridine, even at pressures very near to saturation. Similar experiments with water vapour gave indication of large apparent deviations due, almost entirely, to condensation on the walls of the system. It was of importance to estimate the errors arising from this source under actual experimental conditions.* Before the more accurate experiments with water vapour were undertaken (those of Series II and Series III) a bulb of large capacity was sealed to the burette system in order to reduce the ratio of surface to volume. An experiment was then carried out in which 32 successive quantities of water vapour, after measurement in the burette, were distributed to a small detachable tube containing charcoal, the increase in weight of the latter being noted from time to time. The pressures, both at generation and after distribution, were made to correspond closely to those in an actual extensometric experiment, and ranged as high as 0.8 of saturation towards the end of the run. The gain in weight of the charcoal was 0.0504 gm., as against 0.0499 gm. calculated from the pressure measurements by means of the gas laws. In view of this result we consider the *s*-determinations to be free, in the main, from systematic errors greater than 1 or 2 per cent. In the case of all the vapours studied there is probably some loss of accuracy at pressures greater than 0.9 of saturation.

The influence on the extensometric results of gaseous impurities† contained in the charcoal is discussed in the sections which follow. At the conclusion of each adsorption experiment the purity of the vapour in contact with the charcoal was tested by a determination of the vapour pressures at a series of temperatures down to -78°C . Unless otherwise recorded in what follows this test led to negative results. A possible source of error, which did not lend itself to control, lay in the contamination of the charcoal with mercury vapour; the immersion values of the expansion do not appear to differ appreciably when the charcoal is free from such contamination.

* Such errors would arise from the desorption of water from the walls, caused by the fall of pressure on distribution to the extensometer vessel. The quantity of water required to cover the walls of the measuring system with a layer 25 molecules in thickness, though less than $\frac{1}{2}$ per cent. of that taken by the charcoal at saturation, would represent a substantial fraction of the quantity present in the burette during the Boyle's law tests. Since, at the working temperature, even saturated water vapour shows but slight departure from the gas laws (see, for example, Menzies, 'J. Amer. Chem. Soc.', vol. 43, p. 851 (1921)) the wall effect must be held responsible for the observed deviations. The latter were unimportant at pressures less than 0.6 of saturation.

† Before the experiments were begun, all accessible parts of the apparatus were heated to remove, as far as possible, occluded gases. Most of the experiments described were carried out long after the extensometer had ceased to evolve gas, when heated, at a rate detectable by the McLeod gauge.

3. *Experiments with Benzene.*

Except for two of the experiments of Series III, all runs with benzene were carried out at room temperature.

Series I.—These experiments were carried out before the apparatus assumed the final form shown in fig. 1. In particular, the cut-out A and the connecting tube leading from it to the pumps were absent, so that it was necessary, when subliming the benzene *in vacuo*, to leave the extensometer vessel in communication with the remainder of the system. While this was in progress the extensometer vessel was heated to 300° C., and the heating and pumping continued for some time after the cut-out E, leading to the benzene system, had been closed. Uncontrolled variations amounting sometimes to as much

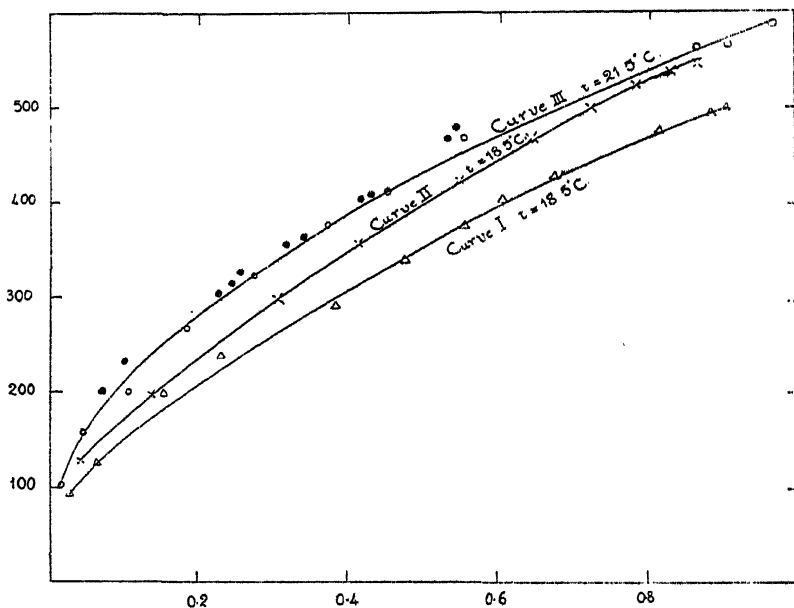


FIG. 2.—Benzene, Series I and II. Abscissæ, values of x ; ordinates, values of Mx/s .

as 30 per cent. of the x value for a given value of s occurred between one experiment and another; extreme values were obtained in the fifth and eighth experiments, the results of which are illustrated in fig. 2, curves I and II. Within certain limits, variations in the temperature of evacuation and of the working temperature were found not to be responsible for these discrepancies. Further, although the extensometer reading for the benzene-free charcoal showed a definite drift—always in the direction of shrinkage—the

discrepancies between the expansion curves followed no definite trend, and could be attributed neither to the gradual removal of traces of a foreign gas nor to a gradual poisoning by the benzene. Only a very small quantity of non-condensable gas was disengaged from the charcoal by the benzene in the course of each experiment.

All other obvious causes having been eliminated, it was concluded that the variations arose through the exposure of the hot charcoal to the gases pumped off the benzene in the course of its sublimation. After the conclusion of the experiments with oxygen-poisoned charcoal, described below, the apparatus was reconstituted in the manner already indicated; not only were the results thereafter in far more satisfactory agreement, but the zero-point now no longer showed the drift to which it had hitherto been subject.

Series II.—With the object of testing directly the effect of foreign gases, a train for the generation of pure dry oxygen was sealed to the pump system. The charcoal was exposed to this gas when hot, and then subjected to prolonged pumping at a lower temperature to remove the loosely held gases; the experiment with benzene was then carried out in the usual way. The quantity of oxygen adsorbed could not be measured.

In a typical experiment the charcoal at 265°C. , was exposed to some 400 mm. pressure of oxygen for 3 hours. The resulting expansion, as measured in the cold, was 0.084 per cent., but prolonged evacuation at 105°C. reduced this to 0.076 per cent., the gas pressure falling meanwhile to rather less than 10^{-5} mm. The results of the subsequent benzene experiment, fig. 2, curve III, leave no doubt as to the effect of oxygen in increasing the expansion for a given quantity of benzene. The rate of adsorption was much reduced. Further, by taking time observations, it was established that the slow adsorption of benzene which took place on standing was accompanied by a smaller expansion than the rapid adsorption which followed an increase of pressure. Thus in fig. 2 the points indicated by black circles, which refer to observations made soon after an admission of vapour, all lie above the curve, curve III, which has been drawn with reference to observations (see plain circles) taken immediately before the pressure was raised. The quantity of non-condensable gas (at -78°C.) disengaged in the course of this experiment was remarkably small, its pressure being of the order of 0.05 mm.

The saturation value of the expansion was but little affected by the pressure of the oxygen; but it is noteworthy that although all other effects of the oxygen poisoning were gradually removed by repeatedly charging the charcoal with benzene and vacuum-heating it, the x - and also probably the s -values at

saturation were *thereafter* somewhat greater. It is possible that the treatment removed some easily oxidisable impurities from the surface.

Series III.—These experiments, which led to highly concordant results, followed a long series with the alcohols, in the course of which the condition of the charcoal and also the experimental procedure and technique had become standardised. They provide a partial explanation of the anomalous behaviour previously described for they place it beyond all doubt that the adsorption of benzene by charcoal II is a complex phenomenon involving two distinct processes. The first of these causes a large expansion of the charcoal and therefore probably amounts to the formation of a two-dimensional gas film. The second, a slower process, is accompanied by a contraction. The behaviour of the benzene was thus similar to that previously described* in the case of pyridine.

The existence of the secondary process is most easily established under the conditions prevailing at low *s*-values, where, some short time after each admission of vapour, the primary adsorption necessarily ceases for lack of further vapour in the deadspace. At this stage—reached, in the case of benzene and pyridine, about half-an-hour after each admission of vapour—the extensometer ceases its forward movement and begins to move in the direction of contraction. If the primary adsorption were also slow, the secondary effect would be detectable only by such a difference in the amount of expansion caused by the rapid and the slow adsorptions as was noted in Series II. The discrepancies in Series I were probably due in part to the secondary process—masked also in this case—taking place to a different extent in each different experiment. It is clear that the presence of impurities—which might affect differentially not only the rates of the primary and secondary processes, but their equilibrium points as well—must be very complex indeed.

Both with benzene and with pyridine the secondary process was found to be accelerated by heating the charcoal. In one of the benzene experiments, for example, the following observations were made :—

Adsorption value.....	0.0825 gm. per gram.
Maximum expansion (2 hours after last admission) ..	0.268 per cent.
Expansion after standing 7 hours at room temperature	0.261 ,,
Expansion after further 17 hours at room temperature	0.258 ,,
Expansion after further 22 hours, during which charcoal was heated for 2 hours to 100° C.	0.243 ,,

* Bangham and Fakhoury, *loc. cit.*, p. 1327.

In fig. 3A, points *a*, *b*, *c*, etc., refer to the maximum observed values of the expansion, and *a'*, *b'*, *c'*, etc., to the "equilibrium" values observed after heating the charcoal to 100° C. and allowing it to cool overnight. At a certain stage in each experiment, roughly coincident with that at which the deadspace pressure became measurable, the tendency to retrograde movement ceased, and, since neither heating to 100° C. nor standing over long periods had any appreciable effect on the reading, the practice of heating the charcoal was discontinued at this stage.

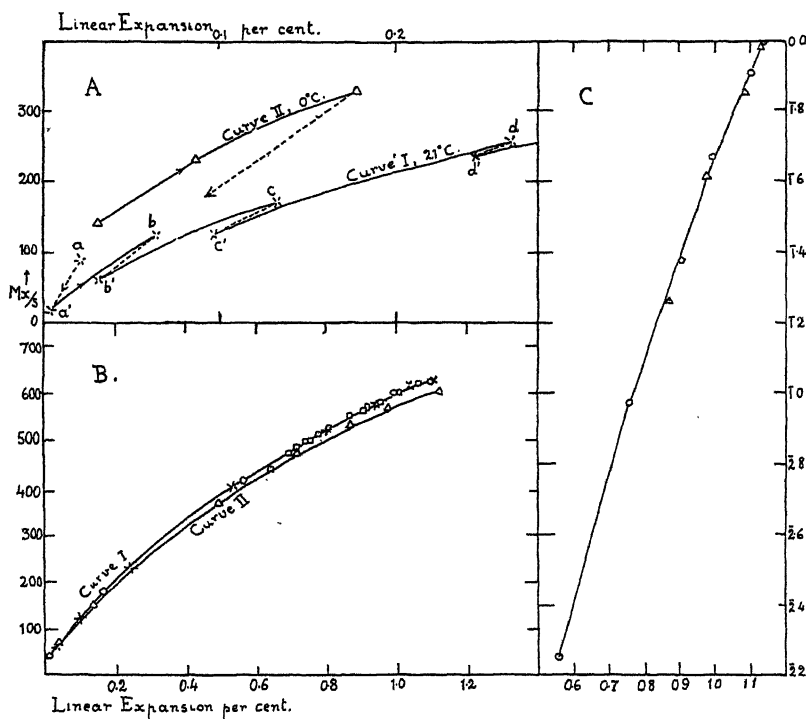


FIG. 3.—A: Molecular expansion curves of benzene. B: "Equilibrium" molecular expansion curves of benzene; \odot adsorption (1) at 22° C.; \times adsorption (2) at 21° C.; \square desorption (2) at 21° C.; \triangle adsorption at 0° C. C: Benzene Series III. Abscissæ values of x ; ordinates, values of $\log p/p_0$; \odot adsorption at 22° C.; \triangle adsorption at 0° C.

The full equilibrium curve for benzene at room temperature is given by curve I in fig. 3B; the graph includes also some points relating to a typical desorption experiment, and illustrates the fact, amply supported by other data, that in this case the molecular expansion curve follows in desorption practically the same course as in adsorption. Although great accuracy

cannot be claimed for points near the axis* it appears reasonable to conclude that at zero adsorption the molecular expansion—if not actually infinitesimal—is a very small fraction of the values (*ca.* 100 per cent., *vide infra*) calculated by equation (6). In general, the course of the equilibrium curve is difficult to explain in terms of liquefaction or intermolecular attraction, particularly in view of the entirely different behaviour of water described in a later section. More probably the secondary process (*i.e.*, that accompanied by a contraction) consists in the greater proportion of the molecules adsorbed at low concentrations becoming attached at favourable points to the charcoal surface. In view of the accelerating influence of heat, it is possible that this process is one which takes place by “activation” either of the adsorbent or of the adsorbate. From the fact, already mentioned, that heating, followed by cooling, was without effect on the readings once the deadspace pressure became considerable, it is to be inferred that the molecules adsorbed between this point and saturation are nearly all mobile.

The above view of the secondary process is to some extent borne out by the results of experiments carried out at 0° C. In one experiment three admissions of vapour were made before the system was heated to bring it to equilibrium. The (non-equilibrium) values of the molecular expansion, see fig. 3A, curve II, are seen to be far higher than those reached at this stage in the experiments at room temperature, and suggest that Mx_0/s_0 is in this case at all events finite. After the third admission of vapour the charcoal was heated to 100° C. before each measurement at 0° C. was made. The equilibrium curve, see fig. 3B, curve II, is seen to lie below that at 21–22° C., as it should if the expansion is due to the thermal motion of the molecules.

Relation between Expansion and Deadspace Pressure.—Since in a normal experiment the adsorption reached nearly two-thirds of its saturation value while the pressure was yet too small for accurate measurement, the isotherms are of little interest. On account of the rapid increase of x , a comparison of the p , x measurements is more illuminating. Since s varies but little over the range in question, a rough test of the validity of equation (4) is readily made by plotting $\log p$ against x . According to the equation the result should be a nearly linear graph, slightly concave to the x -axis, and fig. 3c shows that this is actually the case. For convenience of reference, values of $\log_{10} p/p_0$, where

* Observations subject to an estimated error of more than 10 per cent. have been omitted from the graphs in this paper, the only exception being the point a' of fig. 3A, which has been inserted for general illustrative purposes; the corresponding point has been omitted from the “equilibrium” curve of fig. 3B.

p_0 = saturation pressure, have here been plotted as ordinates. The fact that this device brings into coincidence the data at the two temperatures is possibly without general significance, but it is important to note that the saturation values fall naturally on the curve.* In view of the complex nature of the adsorption, the agreement with (4) can be no more than approximate; but the more or less successful application of equation (1) and of the Gibbs equation is noteworthy. The values of Mx_0/s_0 at 22° C. and 0° C. calculated from the curve by means of equation (6) are 100 and 97 per cent. respectively.

4. Experiments with Pyridine.

Some of the results obtained with this substance have already been described.† On account of the tendency of pyridine to foul the apparatus (particularly the glass-mercury interfaces) only two experiments were attempted; in both of these the measurements were made at room temperature. On reaching saturation in the first of these runs a pressure of about 0.3 mm. of gas non-condensable at -78° C. was found to have developed. This did not occur in the second experiment, but the measurements nevertheless showed good agreement in the two cases. In fig. 4A, which records the molecular expansion data for the second experiment, points *a*, *b* and *c* refer to observations made, respectively, soon after an admission of vapour, after standing 36 hours at room temperature and after heating the charcoal for an hour to 240° C. Except for this occasion the charcoal was maintained at room temperature, the normal interval between successive additions being between 1 and 2 hours; the results, therefore, are not strictly comparable with those for benzene already given.

The graph of $\log p/p_0$ against x for pyridine at 17° C. is shown in fig. 4B. Here also the saturation point ($x = 1.26$) falls naturally on the curve, which, in accordance with the requirements of equation (4), has been drawn slightly concave to the x -axis. The value of Mx_0/s_0 , as calculated by the use of equation (6) is 91 per cent.

5. Experiments with Water Vapour.

Series I.—Before commencing this series it had become necessary to admit air to the extensometer vessel in order to clean the apparatus and purify the

* The x -values under conditions of saturation at 22° C. and 0° C. were 1.136 and 1.140 per cent. respectively. The difference, indistinguishable in fig. 3c, is possibly due to experimental error, but the authors have not at present sufficient data to decide whether the saturation x -values of other substances are as little dependent on the temperature as that of benzene.

† Bangham and Fakhoury, *loc. cit.*

mercury. By repeatedly charging the charcoal with water—each charging being followed by vacuum-heating to 300°C .—a stage was reached at which no measurable quantity of non-condensable gas became disengaged during further experiments. Nevertheless, it is probable that the results of the four experiments of this series were largely affected by the presence of oxygen gases unremoved by this treatment. All four experiments were carried out at room temperature. Like those obtained with charcoal I* under similar conditions, the results were highly irregular in the region of low pressures. The graph

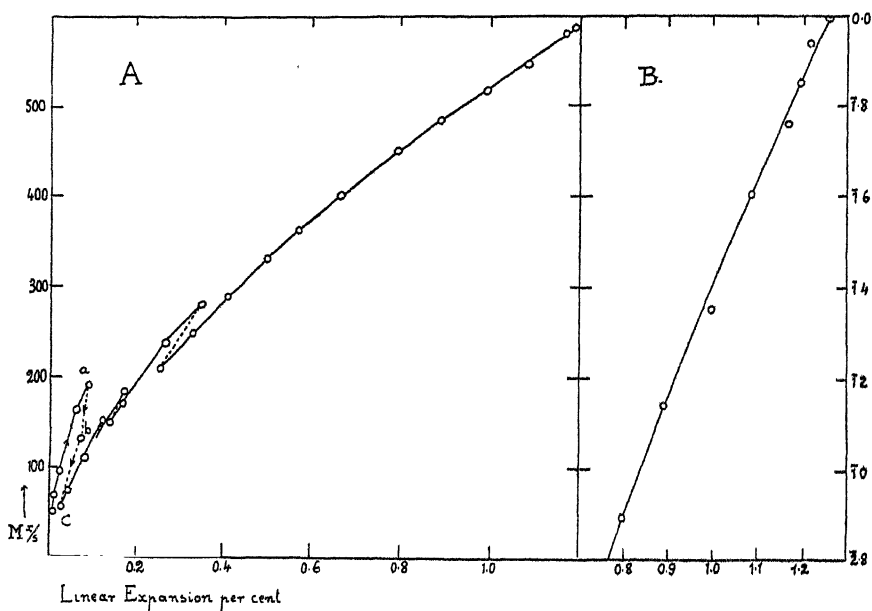


FIG. 4.—A: Molecular expansion curve of pyridine at 17°C . B: Pyridine at 17°C . Abscissæ, values of x ; ordinates, values of $\log p/p_0$.

shown in fig. 5B, which is plotted on a very open scale, is typical of the molecular expansion graphs obtained in this region. Over most of their course the molecular expansion curves are suggestive, in their general trend, of the behaviour of a gas very near its critical temperature, see fig. 5A, curve I. Immediately before saturation, however, they bend again nearly parallel to the abscissa axis. (This bend occurs well beyond the limits of the diagram shown in fig. 5A.)

Series II.—The charcoal (charcoal IIa) used in this and in the next series of experiments was from the same original stock as charcoal II and had been

identically treated. After prolonged exhaustion at 560° C., however, it was saturated with water before being exposed to air. In this condition it was mounted immediately in a new extensometer of magnification ratio 110; the latter was inserted in a tube which was sealed to the apparatus and exhausted as rapidly as possible in the cold. Saturation with water and re-evacuation were repeated. The charcoal was allowed to retain a small residue of water (at a negligible pressure) during the calibration of the tube with hydrogen.*

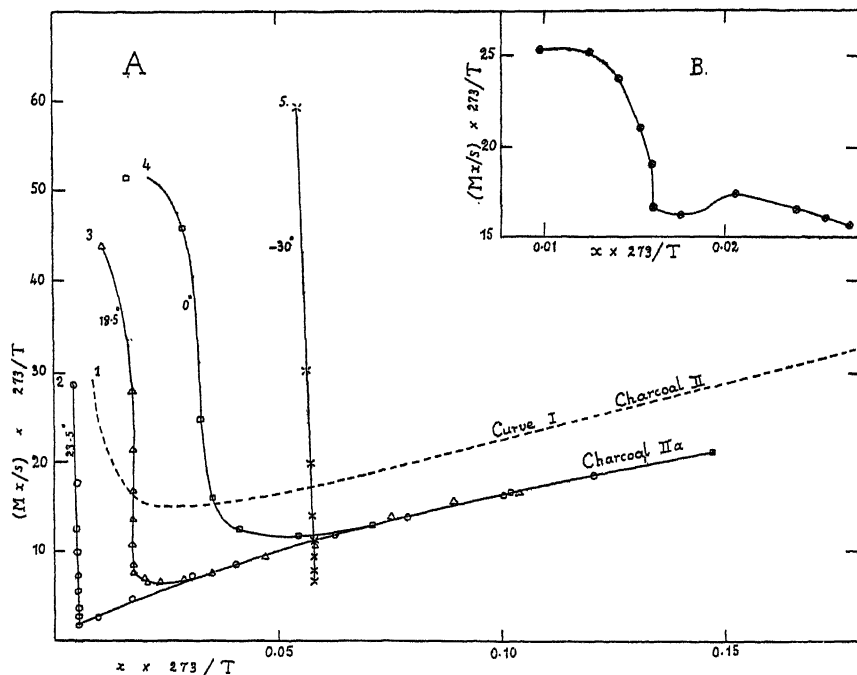


FIG. 5.—Molecular expansion curves of water. A: curve 1, charcoal II at 23.5° C.; curve 2, charcoal IIa at 23.5° C., points \odot ; curve 3, charcoal IIa at 18.5° C., points \triangle ; curve 4, charcoal IIa at 0° C., points \square ; curve 5, charcoal IIa at -30° C., points \times . B: Charcoal II and water at 23.5° C.

After being charged twice again with water, the charcoal was finally heated to 300° C. for several hours.

Five adsorption experiments were then carried out, covering the range of temperatures from -30° to 100° C. In the experiment at 100° C. no measurable

* The intention of this procedure was to make use of an adsorbed film of water to protect the charcoal (after its high temperature treatment) from air and from the hydrogen used in calibration. The extensometer being a (nickel coated) brass "experimental" model, it was impossible to heat the charcoal to 560° C. *in situ*. A further change in the apparatus was made at this stage; a calibrated bulb of large capacity was sealed to the burette system in order to give the latter a larger volume in proportion to its surface.

expansion of charcoal took place, although some 0.0074 gm. of water per gram were taken up. The results of the experiments at -30° , 0° , 18.5° and 23.5° C. are illustrated by the molecular expansion graphs given in fig. 5A; for closer comparison, the values both of the expansion, and of the molecular expansion, have been multiplied by the factor $273/T$, where T is the absolute temperature. The curve for 23.5° C. is similar in form to the pv graph of a gas undergoing condensation by isothermal compression. It is to be noted, however, that while, in the region where the molecular expansion first begins to increase, the curve has a fairly steep slope, which would cause it, if produced back to cut the axis, to do so at the origin, it later bends away from the ordinate axis, becoming ultimately nearly straight and decidedly less steep than curve I for charcoal II. The curves for 18.5° and 0° C. show no break-points, but, after passing through rounded minima, become practically coincident with that for 23.5° C.

The constant λ of equation (1), though possibly a function of the temperature, is hardly likely to be characterised by a large negative temperature coefficient. It is therefore to be inferred, from the relative positions of the molecular expansion graphs at different temperatures (also from the fact that no observable expansion took place at 100° C.) that the water in the adsorbed phase is less condensible the higher the temperature, and the latent heat of two-dimensional vapourisation negative.

The adsorption isotherms at 0° , 18.5° , 23.5° and 100° C. and the desorption isotherms at the two intermediate temperatures are illustrated in fig. 6, where $\log s$ has been plotted against $\log p/p_0$. It is rendered extremely probable by the relative positions of the curves at the different temperatures that, throughout most of the range of s -values investigated, the heat of adsorption from the vapour phase is less than the latent heat of vapourisation of the bulk liquid.* The theory of Semenoff† requires that, in such a case, the fraction

* The excess of the heat of adsorption over the heat of liquefaction is given by

$$RT^2 (\partial \log_e p / \partial T)_s - RT^2 d \log_e p_0 / dT = RT^2 \partial / \partial T (\log_e p / p_0)_s,$$

and is positive or negative according as the isothermals at lower or at higher temperatures lie uppermost in the diagram shown in fig. 6. The evidence is not wholly conclusive, for while the use of the equation presupposes the complete reversibility of the equilibrium with respect to both temperature and pressure, the experimental data show that it is nearly, but not quite, reversible with respect to pressure. In the case of certain other charcoals the heat of adsorption of water at moderate s -values has been found less than the heat of liquefaction; see, for example Coolidge, 'J. Amer. Chem. Soc.,' vol. 49, pp. 708, 1949 (1927); Keyes and Marshall, *ibid.*, p. 156; cf. also, Krut and Modderman, 'Chem. Rev.,' vol. 7, p. 259 (1930). It would have been easy to obtain more decisive evidence as to the behaviour of charcoal IIIa in this respect had not a series of accidents led to the destruction

of the saturation pressure at which the formation of the liquid film sets in should increase as the temperature is lowered. The values in Table I confirm this.

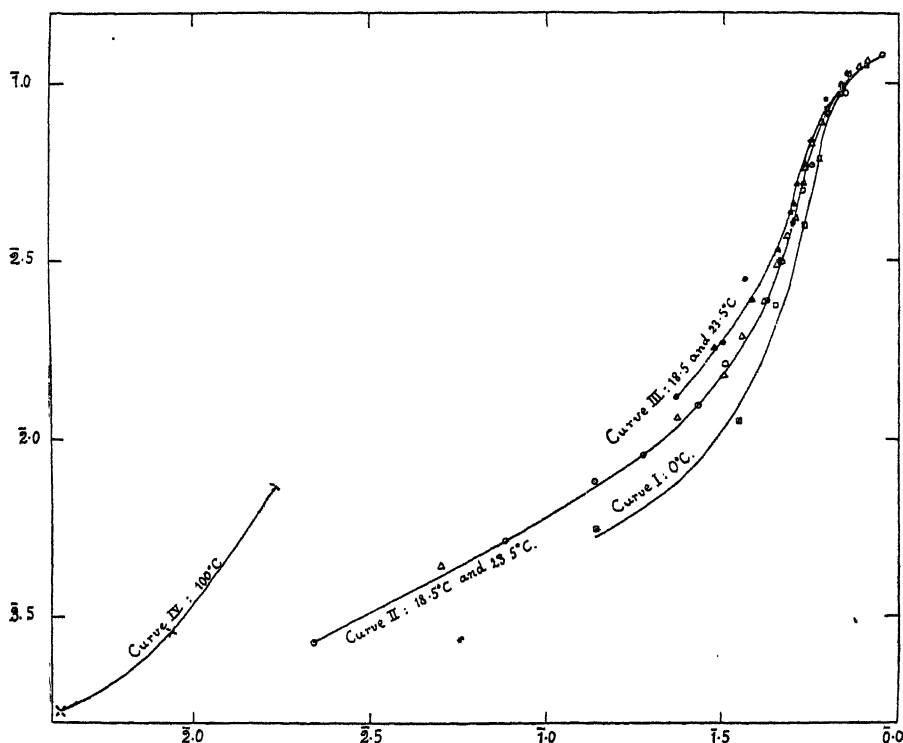


FIG. 6.—Charcoal IIa : Series II, water isotherms. Abscissæ, values of $\log p/p_0$; ordinates of $\log s$ (s in grams). \square Adsorption at 0°C .; \odot adsorption at 18.5°C .; \bullet desorption at 18.5°C .; \triangle adsorption at 23.5°C .; \blacktriangle desorption at 23.5°C .; \times adsorption at 100°C .

of the specimen. The following further points in connection with the curves of figs. 5 and 6 may be noted here : (1) no time-effects were observed, both p and x taking up steady values immediately after each admission or removal of vapour ; (2) the experiment at 23.5°C . having been pushed to saturation, the desorption $\log s$ values for this experiment have been calculated on the basis of water *ultimately* recovered. The desorption values for the 18.5° experiment, on the other hand, are reckoned as differences between the total quantity delivered to the charcoal and the quantities removed therefrom. (In this experiment, in order to minimise the errors, the pressure was not raised to saturation.) The adsorption curves for these two experiments are nearly coincident ; the coincidence of the desorption curves serves to illustrate that, within experimental error, the water adsorbed on charcoal IIa was quantitatively recoverable as such (cf. Allmand, Hand and Manning, 'J. Phys. Chem.,' vol. 33, p. 1694 (1929)). (3) In Series II it was suspected that the molecular expansion curves resulting from desorption measurements followed a course somewhat different from those obtained in the adsorption experiments. The experiments of Series III did not confirm this, however, and the authors prefer to await further evidence before making a statement on this subject.

† 'Z. Phys. Chem.,' B, vol. 7, p. 471 (1930).

Table I.—Values of the Expansion (x), and Minimum Values of the Fraction of Saturation Pressure (p/p_0), at which $dx/ds = 0$.

Temperature.	x .	p/p_0 .
° C.		
100	0.0	0.0
23.5	0.005	<0.077
18.5	0.019	0.24
0.0	0.034	0.45

It is reasonable to suppose (*cf.* Semenoff) that the break-point in the molecular expansion curve at 23.5° C. corresponds to the completion of a unimolecular liquid film. Before this point is reached the surface-liquid and surface-gaseous phases coexist, and the only force effective in the cavities of the charcoal is that due to the surface vapour pressure. As the adsorption proceeds, however, the liquid grows at the expense of the vapour phase, and the complete disappearance of the latter is marked by a sudden expansion due to the thrust of the now continuous film into the cavities of the adsorbent. Though the steady increase of expansion beyond this point must signify a further crowding of molecules into the layer next the charcoal, the increase of adsorption is probably effected more by a thickening of the film than by an increase in concentration in the innermost layer; thus the rising section of the molecular expansion curve is not straight, but bends away from the ordinate axis.

The curve (curve I, fig. 5A, see also fig. 5B) for charcoal II requires a very different interpretation. Whilst some liquefaction of the film is rendered probable by the definite evidence of breaks, fig. 5B, the fact that the rising portion, if produced back to meet the axis, would do so well above the origin, leads to the conclusion that this liquefaction is by no means complete. The saturation value of the expansion was more than double that shown by charcoal IIa at the same temperature.* It is possible that the expansion is due, in part, to the compression by the water film of mobile molecules of foreign gases. The small and constant slope of curve I suggests that, whatever be the phase they form, the water molecules become distributed in some depth at quite an early stage, and have little tendency to spread over the surface as a continuous unimolecular film.

The fact that the curves for charcoal IIa at 18.5° and 0° C., show rounded minima instead of break-points is probably due to the inhomogeneity of the

* 0.40 per cent. as against 0.18 per cent. (charcoal IIa) at 23.5° C.

charcoal, which would make it possible for the water layer to thicken in some regions before it has become continuous in others.* The curve for -30°C . (fig. 5A, curve 5) shows no minimum value but falls vertically, in the later stages of the adsorption, until saturation is reached. It is natural to suppose this to be due to the formation of a *solid* condensed phase.

Series III.—As a check on the results obtained in Series II, charcoal IIa was mounted in the nickel extensometer described in Part I (magnification ratio 40), and the experiments repeated after vacuum-heating the charcoal *in situ* to 460°C . The results, though generally less accurate on account of the lower magnification ratio of the instrument, confirm those already given, and need not be discussed in detail. They do not confirm a suspected departure of the "desorption" molecular expansion curves from those determined in the adsorption experiments.

6. Some Numerical Data : the Specific Surface of Charcoal IIa.

Although results of stoichiometric significance were not aimed at, it is not out of place here to record such of the data as may prove useful for purposes of comparison with later work. In the experiment with water vapour at 23.5°C ., the *s*-values immediately before and immediately after the break in the molecular expansion curve were respectively 0.050 and 0.060 gm. per gram. Assuming that at the actual break-point the charcoal was covered with a layer of liquid water of normal density, and of thickness equal to the cube root of the ratio : molecular volume/Avogadro constant, it is calculated that its area must lie between 160 and 190 square metres per gram.

The highest recorded *s*-value for charcoal IIa, that at 94 per cent. of the saturation pressure (23.5°C .) was 0.123 gm. It appears, therefore, that near saturation the film is between two and three molecules in thickness. The

* Compare the remarks of Semenoff (*loc. cit.*) concerning the masking of the break-points which should occur in the isotherms at adsorption values corresponding to the completion of the first, second, etc., molecular layers of liquid. In both the adsorption and the desorption isotherms resulting from our experiments minor irregularities occur which, since they are greater than could be due to experimental error, are possibly connected with the existence of breaks (*cf.* Allmand and Burrage, 'Proc. Roy. Soc.,' A, vol. 130, p. 610 (1931)). The observations are, however, too widely spaced to do more than define the general trend of the isotherms, and since the exact position of such breaks must remain very doubtful, no attempt is made here to co-ordinate them with the molecular expansion curves. The minimum points on the latter, including the break-point in the curve at 23.5°C ., roughly correspond to the points where the isotherms are steepest, and appear to be associated with points of inflection rather than with breaks.

molecular expansion curve being nearly linear in this region, a limiting adsorption value* can be calculated from its slope; this is found to be 0.174 gm. per gram. Similar calculations based on the results obtained with charcoal II would be clearly unjustifiable in view of the marked curvature of the molecular expansion curves of all three working substances in the region of saturation.†

If it be assumed: (1) that the films of benzene and pyridine are unimolecular, and (2) that their area per gram of charcoal II is equal, or nearly so, to the area of the water films on charcoal IIa, then the area per molecule at saturation is found to lie between 16.5 and 19 Å. (benzene) and 14 and 17 Å. (pyridine). These estimates are given with some reserve, however, for, even granted the truth of the first assumption, the second must rest on the common origin and similarity of treatment of the charcoals, and not on any quantitative evidence as to their behaviour.

Summary.

(1) The description is given of an apparatus suitable for measuring the quantities of volatile liquids adsorbed by a charcoal rod, and for determining the expansion resulting therefrom.

(2) The results of various experiments with a natural pinewood charcoal in contact with the vapours of benzene, pyridine and water are discussed in terms of the hypothesis, suggested in a previous paper, that the expansion is a measure—either exact or approximate—of the pressure in the surface phase. On the basis of this hypothesis it is shown that the Gibbs equation is approximately valid over a considerable range of pressure in the case of benzene and pyridine, but not in that of water vapour.

(3) At low pressures the primary adsorption of benzene and pyridine to form gaseous films is accompanied by a secondary process associated with a contraction of the charcoal.

* Par 1.

† The regularity noted in Part 1, namely that the limiting adsorption values of different substances represented nearly equal volumes of the normal liquids, could not, therefore, be tested. The *saturation* adsorption values of benzene (0.140 gm.) and pyridine (0.166 gm.) can be estimated with some accuracy by extrapolating the molecular expansion curves to the known saturation expansion values—a very small extrapolation in these two cases. As liquids of normal density these quantities would occupy respectively 0.160 and 0.169 c.c. The suggestion that the adsorbed films of these substances are gaseous does not, of course, necessarily conflict with the conclusion, arrived at by Gurvitch and others, that the quantities of different substances adsorbed at saturation depend on their densities in the liquid state.

(4) At temperatures between 0° and 23.5° C. the "molecular expansion" curves for water adsorbed on one of the charcoals used are similar to the pv curves of a vapour undergoing condensation by isothermal compression. The surface phase is, however, more readily condensible at higher than at lower temperatures. At -30° C. the surface vapour phase condenses to form a solid film.

(5) Some effects of the presence of gaseous impurities are described and discussed.

Theory of Uncoupling and Formulæ for the Stark Effect in H_2 .

By J. K. L. MACDONALD, Ph.D., 1851 Exhibitioner, Emmanuel College, Cambridge.

(Communicated by R. H. Fowler, F.R.S.—Received June 8, 1932.)

I. UNCOUPLING.

1.1 *Introduction.*

Kronig and Fujioka† have developed a perturbation theory to account for "uncoupling phenomena" in Hc_2 . These phenomena manifest themselves as a distortion of the usual form of rotational level spacing. In a physical model the effects are attributable mainly to the gyroscopic action in which the rotation of the molecule causes the electronic angular momentum to become uncoupled from the internuclear axis. By this action the component of the electronic angular momentum about this axis does not remain a constant of the motion. Since the work of Kronig and Fujioka was not concerned with the case of H_2 , and since those authors did not discuss the nature and justification of the approximations tacitly employed, the present writer thought it of interest to discuss relevant points and to develop a method for obtaining higher approximations. In a separate communication to this journal Dr. P. M. Davidson is dealing with matters concerning more directly the numerical applications of the theory.

1.2 *The Mathematical Basis.*

Kronig and Van Vleck have shown how to express the wave equation for a diatomic molecule in a convenient way.‡ Kronig's expression is equivalent

† 'Z. Physik,' vol. 63, p. 168 (1930).

‡ "Band Spectra and Molecular Structure," chap. 1, Camb. Univ. Press (1930).

to the following equation, except that our equation includes terms and operator forms not considered by Kronig. The correct wave equation for a diatomic molecule may be written as :

$$[H - W] \Psi = [H_0 + h' - W] \Psi(\xi, \eta, \zeta, \rho_e, \rho, \theta, \psi, s_e, s_n) = 0 \quad (1)$$

In (1),

$$H_0 = \left\{ -\frac{\hbar^2}{8\pi^2 m} (\sum_r \nabla_r^2) + V_0 \right\} - B \left\{ \frac{\partial}{\partial \rho} \left(\rho^2 \frac{\partial}{\partial \rho} \right) + \frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial}{\partial \theta} \right) + \frac{1}{\sin^2 \theta} \left(\frac{\partial}{\partial \psi} - i \cos \theta \mathfrak{L}_z \right)^2 \right\},$$

and

$$h' = B \left\{ \mathfrak{L}_\xi^2 + \mathfrak{L}_\eta^2 - \frac{\partial}{\partial \rho_e} \left(\rho_e^2 \frac{\partial}{\partial \rho_e} \right) - 2\rho^2 \frac{\partial}{\partial \rho} \left(\frac{\partial}{\partial \rho_e} \right) \right\}_{\rho_e = \rho} - \frac{\hbar^2}{8\pi^2 (M_1 + M_2)} \left\{ \sum_{\substack{r, r' \\ r \neq r'}} (\nabla_r \cdot \nabla_{r'}) + V' \right. \\ \left. + B \left\{ i \mathfrak{L}_\xi \left(2 \frac{\partial}{\partial \theta} + \cot \theta \right) + \frac{2i}{\sin \theta} \mathfrak{L}_\eta \frac{\partial}{\partial \psi} + \cot \theta (\mathfrak{L}_\eta \mathfrak{L}_\zeta + \mathfrak{L}_\zeta \mathfrak{L}_\eta) \right\} \right\},$$

$$B = \hbar^2 / 8\pi^2 \mu \rho^2, \quad 1/\mu = 1/M_1 + 1/M_2, \quad 1/m = 1/m_e + 1/(M_1 + M_2).$$

The following symbols have been used: m_e and M_1, M_2 are respectively the electronic and the two nuclear masses. μ and m are "reduced masses." ρ, θ and ψ are spherical polar co-ordinates for the internuclear line.† The electron co-ordinates are measured in the moving, rectangular system ξ, η, ζ with axes determined by the vectors \mathbf{i}, \mathbf{j} and \mathbf{k} along the directions of the arcs $(d\psi), (-d\theta)$ and $(d\rho)$ respectively. The origin is taken at the nuclear centre of mass. We use $\nabla_r = \mathbf{i} \frac{\partial}{\partial \xi_r} + \mathbf{j} \frac{\partial}{\partial \eta_r} + \mathbf{k} \frac{\partial}{\partial \zeta_r}$ for the " r th electron" whose position is given by the vector $\mathbf{r} = \mathbf{i}\xi_r + \mathbf{j}\eta_r + \mathbf{k}\zeta_r$. $\mathfrak{L}_\xi, \mathfrak{L}_\eta$ and \mathfrak{L}_ζ are the ξ, η and ζ components of the electronic angular momentum operator \mathbf{L} , where $i\mathbf{L} = \sum_r [\mathbf{r} \times \nabla_r]$. V_0 contains all the Coulomb potential terms, while V' contains all the spin, internal magnetic and "relativistic" energy terms. In all

† ρ_e is not to be confused with the symbol which has been used formerly for the "equilibrium nuclear separation," now generally denoted by ρ_e . ρ_e is set equal to ρ after the differentiations have been performed. Equation (1) is true for any arbitrary division of the actual dependence of Ψ upon the ρ variable, into dependences upon ρ_e and ρ . In later calculations we shall make this division in a special way so that ρ_e will occur only in a "separated" electronic wave-function which will not be acted on by $(\partial/\partial \rho)$ operators.

subsequent discussion we shall neglect V' since there is strong semi-quantitative support for the view that its energy terms are small in light molecules.

Solution of equation (1).—A separable equation rather similar to (1) is :

$$[H_0 + h'_N - W_N] \Psi_N = 0, \quad (2)$$

where

$$\Psi_N = S_e(s_e) S_n(s_n) \Phi_{g,\Lambda}(\xi, \eta, \zeta, \rho_e) P_{g,\Lambda,K,v}(\rho) \Theta_{K,M,\Lambda}(\theta, \psi)$$

and

$$h'_N = \int \Phi_{g,\Lambda}^* h' \Phi_{g,\Lambda} d\tau_{\xi,\eta,\zeta}.$$

S_e and S_n are electronic- and nuclear-spin functions respectively (see Kronig's book (*loc. cit.*), p. 21). $\Phi_{g,\Lambda}$, $P_{g,\Lambda,K,v}$ and $\Theta_{K,M,\Lambda}$ are respectively the electronic, vibrational and the rotational wave functions (similar to those discussed by Kronig in his book, p. 16) associated with the electron quantum numbers g, Λ , vibration number v and rotation numbers K, M . Collectively g, Λ, v, K, M , and the spins are denoted by N . We may consider W_N as a "zeroth approximation" for the energy W of the correct molecular wave equation (1), and then use the solutions of equation (2) to build up a solution for (1). $\Phi_{g,\Lambda}$ satisfies the "fixed nuclei" electron wave equation :

$$\left\{ \frac{\hbar^2}{8\pi^2 m} (\sum_r \nabla_r^2) - V_0 + W_{g,\Lambda}(\rho_e) \right\} \Phi_{g,\Lambda}(\xi, \eta, \zeta, \rho_e) = 0. \quad (3)$$

$W_{g,\Lambda}(\rho_e)$ may be called the electronic energy for a nuclear separation equal to ρ_e . Due to symmetry about the internuclear line $\Phi_{g,\Lambda}$ depends upon a co-ordinate ϕ only in a factor $e^{i\Lambda\phi}$, ϕ being the angle specifying the orientation of the electron system about the internuclear axis. Therefore, Λ measures the angular momentum of the electron system about this axis, *i.e.*, $\mathfrak{L}_\zeta \Phi_{g,\Lambda} = \Lambda \Phi_{g,\Lambda}$. Equation (3) could not be separated out from the molecular equation (1) since (1) contains gyroscopic terms (in h') which, in (2), are averaged over the electron motion given by (3). Simple investigation of the forms h' and $\Phi_{g,\Lambda}$ shows that h'_N has the form

$$h'_N = \int \Phi_{g,\Lambda}^* \left[B \left\{ \mathfrak{L}_\xi^2 + \mathfrak{L}_\eta - \frac{\partial}{\partial \rho_e} \left(\rho_e^2 \frac{\partial}{\partial \rho_e} \right) \right\} - \frac{\hbar^2}{8\pi^2 (M_1 + M_2)} \left\{ \sum_{r \neq r'} (\nabla_r \cdot \nabla_{r'}) \right\} \right] \Phi_{g,\Lambda} d\tau_{\xi,\eta,\zeta} = h'_N(\rho_e).$$

$\Theta_{K,M,\Lambda}$ satisfies the "symmetrical top" wave equation† :

$$\left\{ \frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial}{\partial \theta} \right) + \frac{1}{\sin^2 \theta} \left(\frac{\partial}{\partial \psi} - i\Lambda \cos \theta \right)^2 + K(K+1) - \Lambda^2 \right\} \Theta_{K,M,\Lambda}(\theta, \psi) = 0 \quad (4)$$

† 'Z. Physik,' vol. 39, p. 444 (1926); 'Phys. Rev.,' vol. 33, p. 476 (1929).

$$\Theta_{K, M, \Lambda} = \left[\frac{(d+s+p)! (1+d+s+2p) e^{2iM\psi}}{p! (d+p)! (s+p)! 2\pi t^d (1-t)^s} \right]^{\frac{1}{2}} \left(\frac{\partial}{\partial t} \right)^p [t^{d+p} (1-t)^{s+p}],$$

$$t = \frac{1}{2} (1 - \cos \theta), \quad s = |M + \Lambda|, \quad d = |M - \Lambda|, \quad p = K - \frac{1}{2} (s + d).$$

K measures the total angular momentum of the molecule, $K \geq |M|, |\Lambda|$. $P_{q, \Lambda, K, v}(\rho)$ satisfies the "nuclear vibration" wave equation:

$$\left\{ B \frac{\partial}{\partial \rho} \left(\rho^2 \frac{\partial}{\partial \rho} \right) - W_{q, \Lambda} - h'_N(\rho) + B\Lambda^2 - BK(K+1) + W_N \right\} P_{q, \Lambda, K, v}(\rho) = 0. \quad (5)$$

If one knows the form of the "nuclear potential"

$$U_n(\rho) = [W_{q, \Lambda}(\rho) + h'_N(\rho) - B\Lambda^2]$$

in (5), then one may be able to solve (5) to obtain W_N . In most cases U_n is not known theoretically with even a moderate degree of accuracy. Therefore, in practice, some sort of power series expansion is assumed to be valid in the important range of ρ -values, the coefficients being obtained from calculations using the experimentally determined energies. In the Kratzer-Fues† expression we have:

$$U_n(\rho) = E_0 + U(\rho),$$

where E_0 is a constant which is formally termed the "electronic energy" and $U(\rho)$ is as given below:

$$U(\rho) = (2\pi\nu_0)^2 \mu \rho_0^2 \left[\frac{1}{2} - \left(\frac{\rho_0}{\rho} \right) + \frac{1}{2} \left(\frac{\rho_0}{\rho} \right)^2 - c_3 \left(\frac{\rho - \rho_0}{\rho_0} \right)^3 - c_4 \left(\frac{\rho - \rho_0}{\rho_0} \right)^4 + \dots \right], \quad (6)$$

where ρ_0 is the value of ρ at the "equilibrium position" of the nuclei, where $U(\rho)$ has a minimum value zero. ν_0 is the classical frequency of small oscillations of the nuclei about ρ_0 . For this function we have:

$$\begin{aligned} W_N = E_0 - B_0 \left(\frac{7}{4} c_3 + \frac{7}{8} c_3^2 + \frac{3}{4} c_4 \right) + h\nu_0 \left(v + \frac{1}{2} \right) \left[1 - \frac{3}{2} \gamma^2 (1 + 2c_3) (K + \frac{1}{2})^2 \right] \\ + B_0 (K + \frac{1}{2}) \left[1 - \gamma^2 (K + \frac{1}{2})^2 \right] \\ - B_0 \left(v + \frac{1}{2} \right)^2 (3 + 15c_3 + \frac{1}{2} c_3^2 + 3c_4) + \dots, \\ B_0 = h^2 / 8\pi^2 \mu \rho_0^2, \quad \gamma = 2B_0 / h\nu_0. \quad (7) \end{aligned}$$

Many of the observed levels fit well into the form (7). We shall now consider cases where there are marked departures from this form.

Methods for Successive Approximations.—Any observed departure of the observed energies W from the form (7) is to be attributed to the difference between the operators of (1) and (2). We proceed to build our solutions of

† 'Ann. Physik,' vol. 80, p. 367 (1926).

(1) by a variation process involving a series of Ψ_N functions. [The variation method is very efficient in that it produces "automatically" the closest approximations to the correct "Eigenwerte," subject to the limitations of the available functions. It also makes the problem mathematically definite, avoiding the use of irrelevant unknown terms which are set equal to zero in the late states of the usual extended perturbation theory.] Before we start we should note that (1) is invariant for interchange of the positional co-ordinates of the nuclei, and therefore Ψ is either symmetrical or anti-symmetrical for interchanges of this nature. We therefore form Ψ_N functions to have the desired symmetry properties (using the degeneracy associated with $\pm \Lambda$ for the cases where $\Lambda \neq 0$). To do this we use the form of Ψ_N (given below equation (2)) when $\Lambda = 0$, and the following form when $\Lambda \neq 0$ † :

$$\Psi_{N_{\Lambda \neq 0}} = 2^{-\frac{1}{2}} S_s(s_s) S_n(s_n) P_{q, \Lambda, K, v}(\rho) [\Phi_{q, v}(\xi, \eta, \zeta, \rho_s) \Theta_{K, M, \Lambda}(\theta, \psi) \pm \Phi_{q, -\Lambda} \Theta_{K, M, -\Lambda}]. \quad (8)$$

We now apply the variation process, thus :

$$\Psi = \sum_N C_N \Psi_N \text{ subject to } \int \Psi^* \Psi d\tau = 1 \text{ and } \delta \int \Psi^* [H_0 + h' - W] \Psi d\tau = 0. \quad (9)$$

The C_N 's are variation parameters and the integration extends over the complete configuration space. The relations (9) determine the values of W and the C_N 's. Using equations (2) and (9) we have :

$$\sum_{N_s} [(W - W_{N_r}) \delta(N_r, N_s) - \omega_{N_r N_s}] C_{N_s} = 0 \text{ for all } N_r \text{'s}. \quad (10)$$

Here we have used the normal, orthogonal properties of the Ψ_N 's and the symbol

$$\omega_{N_r N_s} = \left[\int \Psi_{N_r}^* (h' - h'_{N_s}) \Psi_{N_s} d\tau \right] \quad (11)$$

(note that $\omega_{N_r N_r} = 0$).

We now let English subscripts r, s, \dots , refer to N_r, N_s, \dots , for a group of states near the level which we are investigating, states for which $1/(W - W_s)$ and (or) $\omega_{rs}/(W - W_r)$ will be found to be comparatively large. We let Greek subscripts σ, τ, \dots , refer to N_σ, N_τ, \dots , for the other states used in the variation function. This grouping of the states will enable us to avoid con-

† States are termed $\Sigma, \Pi, \Delta, \dots$, according as $\Lambda = 0, 1, 2, \dots$. For $\Lambda \neq 0$ we have "a" and "b" states associated with the (+) and (-) in (8), e.g., Π_a, Π_b .

vergence difficulties which may occur in the usual perturbation series expressions. Equations (10) then become :

$$(W - W_r) C_r - \sum_s \omega_{rs} C_s = \sum_\sigma \omega_{r\sigma} C_\sigma \quad \text{for all } r\text{'s}, \quad (12)$$

$$C_\sigma = \sum_s \Omega_{\sigma s} C_s + \sum_\tau \Omega_{\sigma\tau} C_\tau \quad \text{for all } \sigma\text{'s}, \quad (13)$$

where

$$\Omega_{\sigma s} = \omega_{\sigma s} / (W - W_\sigma).$$

Successive applications of (13) give

$$C_\sigma = \sum_s \left\{ \Omega_{\sigma s} + \sum_{\tau_1} \Omega_{\sigma\tau_1} \Omega_{\tau_1 s} + \sum_{\tau_1, \tau_2} \Omega_{\sigma\tau_1} \Omega_{\tau_1\tau_2} \Omega_{\tau_2 s} + \dots \right\} C_s \quad (14)$$

(supposed convergent or asymptotic). Equation (12) then may be written as

$$[W - W_r - \varepsilon_{rr}] C_r - \sum_s [\omega_{rs} + \varepsilon_{rs}] C_s = 0 \quad \text{for all } r\text{'s}, \quad (15)$$

where

$$\varepsilon_{rs} = \sum_\sigma \omega_{r\sigma} \left\{ \Omega_{\sigma s} + \sum_{\tau_1} \Omega_{\sigma\tau_1} \Omega_{\tau_1 s} + \sum_{\tau_1, \tau_2} \Omega_{\sigma\tau_1} \Omega_{\tau_1\tau_2} \Omega_{\tau_2 s} + \dots \right\}. \quad (16)$$

Due to the manner in which we have chosen our "Greek subscript" states the Ω 's in (16) are supposed to be small, hence ε_{rs} is presumably small (as compared with the neighbouring terms in (15)). We proceed therefore as follows in our successive approximating :—

- (i) Neglect the ε 's in (15) and find W values in the r, s, \dots states by using the resulting eliminant of the linear equations of (15).
- (ii) Use the W 's obtained in (i) to obtain values for the ε 's which are then used in (15) to give corrected values for the W 's (as determined from the new eliminant).
- (iii) Repeat the process (ii) with the newly found W 's and continue the process until the apparent convergence is sufficient.

If it is found that a few of the "Greek subscript" states cause convergence difficulties they should be treated as "English subscript" states.

The above method is applicable if one knows the $W_r, W_s, \dots, W_\sigma, W_\tau, \dots$ values accurately from a theoretical calculation. In practice, however, we know the values of the ω 's to a fair degree of approximation, and we know the experimental W values. From this knowledge we wish to determine the W_r, W_s, \dots values and the coefficients C_N (in order to calculate the intensities). For this purpose we may proceed as follows :—

- (i) Neglect the ε 's in (15) and use the experimental W -values to determine the W_r, \dots (and also the W_σ, \dots from a similar problem for these energies).

- (ii) Use the W_σ, \dots , values of (i) to find the Ω 's, whence the ε 's. Substitute the W_r, W_s, \dots , and ε_{rs} values in (15) and obtain a new set of W_r, \dots , and W_σ, \dots .
- (iii) Repeat the process (ii) with the new set of $W_\sigma, \dots, W_r, \dots$, and continue the process until the apparent convergence is satisfactory.

Alternatively, since $W - W_\sigma$ is supposed to be comparatively large for the W 's associated with the r, s, \dots , states, we may use the experimental W -values for the σ, τ, \dots , states instead of the W_σ, \dots , in calculating the ε -values. This will probably give a very good approximation for the W_r, \dots , calculated from (15).

To calculate the C_N 's, use the equations which are finally taken to represent (14) and (15), together with the normalising condition in (9) which gives :

$$1 = \sum_N |C_N|^2. \quad (17)$$

Properties of ω 's.—Using the symmetry properties of the Ψ_N 's and of h' , together with the properties of the rotation function $\Theta_{K, M, \Lambda}$, we have $\omega_{N_r N_s} = 0$, unless the following relations hold :—

- (1) N_r and N_s refer to differing q, Λ 's (*i.e.*, differing electronic states).
- (2) $K_r = K_s, M_r = M_s$, and N_r and N_s refer to the same type of member of the spin multiplets (as in the case of transitions).
- (3) Ψ_{N_r} and Ψ_{N_s} have the same nuclear interchange symmetry, *i.e.*, both are “ s ” or “ a .”
- (4) The $\Phi_{q\Lambda}$'s corresponding to N_r and N_s must have the same electron symmetries for reflection of the electron co-ordinates through the mid-point on the internuclear axis (in homonuclear molecules), *i.e.*, both are “ g ” or “ u .” This condition requires that $|L_r - L_s|$ be zero or an even integer, L being the quantum number for the total angular momentum in the corresponding “united nuclei” atomic state.
- (5) $\Lambda_r = \Lambda_s$ or $\Lambda_r \pm 1 = \Lambda_s$. For the case $\Lambda_r \pm 1 = \Lambda_s$ and with appropriate normalisation phases, we have, using the integrals evaluated by Van Vleck† :

$$\omega_{N_r N_s} = x \sqrt{\{(K_r \mp \Lambda_r)(K_r \pm \Lambda_r + 1)\}} \int \Phi_{q_r \Lambda_r}^* P_r 2B \mathfrak{A}_\eta \Phi_{q_s \Lambda_s} P_s d\tau_{\xi\eta\zeta} \rho^2 d\rho, \quad (18)$$

† ‘Phys. Rev.’ vol. 33, p. 479 (1929).

where

$$x = 1 \quad \text{if } \Lambda_r \neq 0 \neq \Lambda_s, \quad \text{and } x = \sqrt{2} \quad \text{otherwise; } P_r \equiv P_{q_r \dots v_r}(\rho).$$

Use of Approximate Electronic Wave Functions.—The writer has discussed elsewhere† an example of the following type of approximate electronic wave function‡ for H_2 :

$$\Phi_{nL\Lambda}(a, b, \rho_s) = [2(1 + \beta S_{nL\Lambda}^2)]^{-1} [M_+(a, \rho_s)(nL\Lambda | b) + \beta M_+(b, \rho_s)(nL\Lambda | a)],$$

$$\beta = +1 \quad \text{for singlets} \quad \beta = -1 \quad \text{for triplets} \quad (19)$$

$M_+(a, \rho_s)$ is an axially symmetrical “inner electron” function (*e.g.*, that of the normal H_2^+ state) for electron “a.” The expression

$$(nL\Lambda | b) = N(n, L, \Lambda) R_{nL}(r_b) P_L^{\Lambda}(\cos \theta_b) e^{i\Lambda\phi_b}$$

is the “outer electron” central field wave function of electron “b,” expressed in polar co-ordinates referred to the mid-point of the internuclear axis. $M_+(a, \rho_s)$, $(nL\Lambda | b)$ and $\Phi_{nL\Lambda}$ are normalised and

$$S_{nL\Lambda}(\rho_s) = \int M_+(a, \rho_s)(nL\Lambda | a) r_a^2 dr_a \sin \theta_a d\theta_a d\phi_a. \quad (20)$$

The wave functions (19) have been found to give “variation energies” in quite good agreement with the observed values of the H_2 energies for the higher excited electron levels. We may expect that these $\Phi_{nL\Lambda}$ ’s will give approximately correct results when used to calculate the ω ’s. This expectation finds support when we build up a variation series $\Phi_{q\Lambda} = \sum_{L=0}^{n-1} C_{nL\Lambda} \Phi_{nL\Lambda}$.

In the cases investigated ($n \leq 3$) it was found that in such series one of the coefficients $C_{nL\Lambda}$ was of an order 10^6 times any other $C_{nL\Lambda}$. Thus the corresponding $\Phi_{nL\Lambda}$ would seem to be a reasonable approximation to the actual $\Phi_{q\Lambda}$. Physical considerations also support this view, since the more highly excited “outer electron” states will be associated with the outer electron far removed from the inner H_2^+ -like structure, and moving in a nearly central field. We should expect, therefore, that integrals relating to angular momentum of the electrons in particular, should be represented quite accurately by calculations using these “central field” wave functions when $L \neq 0$. If we use the $\Phi_{nL\Lambda}$ ’s we find:

$$\text{Case } \Lambda_r \pm 1 = \Lambda_s \quad (\Lambda_r = \Lambda).$$

† ‘Proc. Roy. Soc.,’ A, vol. 136, p. 582 (1932).

‡ The function is appropriate for internuclear separation near ρ_0 .

Using equations (18) and (20) we have :

$$\begin{aligned}\omega_{N_r N_s} = x \sqrt{\{(K \mp \Lambda)(K \pm \Lambda + 1)\}} \int_0^\infty [\sqrt{\{(L_r \mp \Lambda)(L_r \pm \Lambda + 1)\}} \\ \times \{(1 + \beta S_r^2)(1 + \beta S_s^2)\}^{-\frac{1}{2}} \delta(L_r, L_s) \delta(n_r, n_s) \\ + \{\beta T_{rs} T_{rs}\} \delta(\Lambda_r \pm 1, \frac{1}{2} \pm \frac{1}{2})] BP_r P_s \rho^2 d\rho.\end{aligned}$$

We have used

$$S_r = S_{n_r L_r \Lambda_r}, T_{rs} = S_r (1 + \beta S_s^2)^{-\frac{1}{2}} = T_{rs}(\rho_s)$$

and

$$\beta = +1 \quad \text{for singlets,} \quad \beta = -1 \quad \text{for triplets.}$$

Case $\Lambda_r = \Lambda_s (= \Lambda)$.

$$\begin{aligned}\omega_{N_r N_s} = \beta \int_0^\infty \left[P_r B \left\{ T_{rs} (L_r(L_r + 1) + L_s(L_s + 1) - \Lambda^2 \right. \right. \\ \left. \left. - \frac{\partial}{\partial \rho_s} \left(\rho_s^2 \frac{\partial}{\partial \rho_s} \right) - 2\rho^2 \frac{\partial}{\partial \rho} \left(\frac{\partial}{\partial \rho_s} \right) \right) \right]_{\rho_s = \rho} T_{rs} \delta(\Lambda, 0) \\ \left. - (S_r S_s \delta(\Lambda, 1) - F_{rs}) \right\} P_s \int \rho^2 d\rho ; \\ F_{rs} = \frac{\hbar^2}{8\pi^2 (1846) 2m_e} \int \Phi_r^* [\nabla_a \cdot \nabla_b + \nabla_b \cdot \nabla_a] \Phi_s d\tau_{ab}.\end{aligned}$$

Orders of Magnitudes.—(a) S_{nLA} and F_{rs} : $S_{nLA} = 0$ unless $\Lambda = 0$ and L is zero or an even integer. For a nuclear separation $\rho_e = 2$ atomic radii, $S_{320}^2 \sim 0.0001$, $S_{300}^2 \sim 0.001$, and $S_{200}^2 \sim 0.006$. For larger n 's, $S_r^2 \sim (10n_r)^{-3}$, the order of magnitude decreasing very rapidly as L increases, for a fixed value of n . The values for other internuclear separations and the derivatives with respect to ρ_e are correspondingly small. In all cases, $S_r \doteq T_{rs}$.

F_{rs} is calculable for any case, but the resulting form is rather complicated. For our purposes we need note only that $F_{rs} = 0$ unless L_r and L_s are odd and unless $\Lambda_r = 0$ or 1, and that it is of the order 0.0001 for integrals in which the $2p$ states correspond to one of the N 's.

(b) The integral

$$B_{rs} = \int_0^\infty BP_r(\rho) P_s(\rho) \rho^2 d\rho, \quad B = (\hbar^2/8\pi^2 \mu \rho^2).$$

If we use the Kratzer-Fues form (6), with $C_3 = C_4 = \dots = 0$, then

$$P_r(\rho) = \left[\left(\frac{2}{\lambda_r} \right)^{\nu_r+2} \frac{\nu_r! e^{-2\pi i \alpha_r}}{2\sigma_r \Gamma(\nu_r + \alpha_r + 1)} \right]^{\frac{1}{2}} \rho^{(\alpha_r-1)/2} e^{-\rho/\lambda_r} \frac{L_{\nu_r+\alpha_r}^{\alpha_r}(2\rho/\lambda_r)}{\Gamma(\nu_r + \alpha_r + 1)},$$

where

$$\alpha_r = 2 [\gamma_r^{-2} + (K + \frac{1}{2})^2]^{\frac{1}{2}}, \quad \gamma_r = (2B_0/\hbar\nu_0)_r (\doteq 0.05 \text{ in } H_2),$$

$$\sigma_r = v_r + \frac{\alpha_r + 1}{2}, \quad \lambda_r = \sigma_r (\rho_0)_r \gamma_r^2.$$

We have

$$B_{rs} = \bar{B}_0 k_{rs} \times [\text{coefficient of } x^{v_r} y^{v_s} \text{ in } I_{rs}(x, y)],$$

where

$$\bar{B}_0 = [\hbar^2/8\pi^2\mu (\rho_0)_r (\rho_0)_s],$$

$$k_{rs} = (\rho_0)_r (\rho_0)_s \Gamma\left(\frac{\alpha_r + \alpha_s}{2}\right)$$

$$\times \left[\left(\frac{2}{\lambda_r}\right)^{\alpha_r+2} \left(\frac{2}{\lambda_s}\right)^{\alpha_s+2} \left(\frac{\lambda_r \lambda_s}{\lambda_r + \lambda_s}\right)^{\alpha_r + \alpha_s} \frac{v_r! v_s!}{4\sigma_r \sigma_s \Gamma(v_r + \alpha_r + 1) \Gamma(v_s + \alpha_s + 1)} \right]^{\frac{1}{2}}$$

$$I_{rs}(x, y) = (1+x)^{\alpha_r + v_r} (1+y)^{\alpha_s + v_s} [1 + 2(\lambda_s x + \lambda_r y)/(\lambda_r + \lambda_s)]^{-(\alpha_r + \alpha_s)/2}$$

or

$$(1-x)^{-1-\epsilon} (1-y)^{-1+\epsilon} \left[1 - xy - \frac{\lambda_r - \lambda_s}{\lambda_r + \lambda_s} (x-y) \right]^{-(\alpha_r + \alpha_s)/2}, \quad \epsilon = (\alpha_r - \alpha_s)/2.$$

When the vibration numbers v_r, v_s are small as compared with the α 's, and if $\alpha_r \doteq \alpha_s = \alpha$ ($\alpha \doteq 40$ in most of the H_2 states), then,

$$B_{rs} = \bar{B}_0 [(g! / l!) \Gamma(\alpha + 1 + l) / \Gamma(\alpha + 1 + g)]^{\frac{1}{2}} [1 + O(1/\alpha)],$$

where l is the lesser g and the greater of the numbers v_r and v_s . Thus,

$$B_{rs_{v_r=v_s}} \doteq \bar{B}_0, \quad \text{while} \quad B_{rs_{v_r \neq v_s}} \doteq \bar{B}_0 \left[\frac{g}{g + \alpha + 1} \right]^{(g-l)/2}.$$

Similar results have been obtained for other types of vibration functions by Davidson (*loc. cit.*). We shall discuss a different method for dealing with the ρ -variable in a later section (1.4).

It may be seen from the above discussion on the orders of magnitudes that the largest ω -values are associated with the relations

$$(K, M, n, L, v)_r = (K, M, n, L, v)_s, \quad \Lambda_r \pm 1 = \Lambda_s, \quad (\text{spins})_r \equiv (\text{spins})_s,$$

for which

$$\omega_{rs} \doteq x[(K \mp \Lambda_r)(K \pm \Lambda_r + 1)(L \mp \Lambda_r)(L \pm \Lambda_r + 1)]^{\frac{1}{2}} B_{rs}, \quad B_{rs} \doteq \bar{B}_0 \quad (21)$$

All other types of ω 's involve small factors such as S^2 and $(1/\alpha)$, (or zero).

It should be noted that our $\Phi_{nL\Lambda}$ functions refer to levels associated with an unexcited "inner electron." It is possible that certain levels, reported by

Richardson,[†] may be related to states of atomic helium in which *both* electrons are excited. Rough calculations by the present writer indicate that ω 's associated with these types of states may be small as compared with the type of ω which we have just been considering. However, it is probable that in certain cases the former type of ω will be of importance.

1.3. The d^1 -states in H_2 .

By use of the properties (1) to (5) of the preceding section it will be seen that d^1 -states give non-zero ω 's in conjunction with s^1 , d^1 , g^1 , ..., states with the same K and M values and with the same symmetries. In H_2 the set $nd^1 \{\Sigma, \Pi_a, \Pi_b, \Delta_a, \Delta_b\}$, for a given (n, K, v) , is compact and comparatively far separated from the other s^1 , d^1 , g^1 , ..., states (having an unexcited "inner electron"). Moreover, we saw above that ω_{rs} is comparatively large for $(K, M, n, L, v)_r = (K, M, n, L, v)_s$, $\Lambda_r \pm 1 = \Lambda_s$, and for the same symmetry properties for Ψ_{N_r} and Ψ_{N_s} . Therefore we may expect that the first approximation (i), p. 188, will be good if we consider ω 's within the above nd^1 set (n, K, M, v) constant. For corresponding K 's: Σ , Π_a , Δ_a have one type of symmetry while the pair Π_b , Δ_b has the opposite type of symmetry (for interchange of the nuclei). This leads to the following simple determinants for finding the energies:

$$\begin{vmatrix} W - W_0 & -\omega_{01} & 0 \\ -\omega_{10} & W - W_1 & -\omega_{21} \\ 0 & -\omega_{21} & W - W_2 \end{vmatrix} = 0 \text{ for } nd^1 \{\Sigma, \Pi_a, \Delta_a\}_{K, M, v}. \quad (22)$$

$$\begin{vmatrix} W - W_1 & -\omega_{12} \\ -\omega_{21} & W - W_2 \end{vmatrix} = 0 \text{ for } nd^1 \{\Pi_b, \Delta_b\}_{K, M, v}. \quad (23)$$

We have used the subscripts $_0$, $_1$ and $_2$ to refer to the numbers $\Lambda = 0, 1$ and 2 respectively, *e.g.*, $W_0 = W_{N_0}$, $N_0 \equiv \{n, L = 2, \Lambda = 0, v, K, M, \text{spins}\}$.

Using the form (21) we have:

$$\omega_{01} = \omega_{10} = 2\sqrt{3} B_{01} \sqrt{K(K+1)}; \quad \omega_{12} = \omega_{21} = 2B_{12} \sqrt{(K-1)(K+2)}.$$

These determinants for H_2 are equivalent to those used by Kronig and Fujioka (*loc. cit.*) for He_2 .

We may obtain higher approximations corresponding to (ii) and (iii) on p. 189. In many cases such improvements on the forms (22) and (23) modify

[†] 'Proc. Roy. Soc.,' A, vol. 126, p. 500 (1930).

the energy values but slightly. (For some levels there seem to be "extra" perturbations apparently due to some unknown levels. There is some confusion in the analysis of the triplet d -states, and the present uncoupling theory cannot be applied yet to these levels.)

Intensities.—The d^1 -states combine with the $2p^1\Sigma$ and the $2p^1\Pi$ states. It is of interest to calculate the intensities. The $2p^1$ states exhibit little uncoupling, apparently due to their isolation. Therefore, we may use the simple form Ψ_N for the $2p^1$ wave functions. The d^1 -states, on the other hand, have a series form, e.g.,

$$\Psi_{d^1\pi} = S_{00}\Psi_0 + S_{01}\Psi_1 + S_{02}\Psi_2, \quad (\text{or } \Psi_{d^1\pi} = (S_{11b}\Psi_1 + S_{12b}\Psi_2)).$$

The coefficients S_{00} , S_{01} , S_{02} can be determined from

$$\begin{aligned} [(W - W_0)S_{00} - \omega_{01}S_{01}] &= 0; \quad [-\omega_{12}S_{01} + (W - W_2)S_{02}] = 0; \\ [S_{00}^2 + S_{01}^2 + S_{02}^2] &= 1, \end{aligned}$$

where $(W - W_0)$ and $(W - W_1)$ (associated with the W for the $d^1\Sigma$ state under consideration) are found from the determinant (22). (See Kronig and Fujioka, *loc. cit.*, for the same type of formulæ for He_2 .) The intensity for a transition $\Psi' \rightarrow \Psi''$ (e.g., $\Psi_{3d^1\pi, K, M, v} \rightarrow \Psi_{2p^1\pi, K-1, M, v+1}$) is proportional to

$$I = \nu^4 \left| \int \mathbf{V} (\Psi'^* \Psi'' + \Psi''^* \Psi') d\tau \right|^2. \quad (24)$$

In (24), ν is the frequency of the line produced in the transition. \mathbf{V} is the electric vector for all the particles in the molecule. For the usual, unpolarised lines (in the absence of external fields) these intensities are averaged over the M transitions in the manner discussed by Rademacher and Reiche.[†] Using our Φ_{nLA} functions and neglecting terms of the order S_{nLA}^2 , this average intensity is proportional to \bar{I} , where (for $d^1\Sigma \rightarrow 2p^1\Sigma$, say):

$$\bar{I} = \nu^4 [S_{00}A_0B_0 + S_{01}A_1B_1 + S_{02}A_2B_2]^2. \quad (25)$$

$B_r = \int_0^\infty P_{rv}(\rho) P''(\rho) \rho^2 d\rho$, where $P_{rv}(\rho)$ is the vibration function in the Ψ_N , associated with the d^1 -state having $\Lambda = r$ and $v_r = v$; $P''(\rho)$ is the vibration function for the final state. These vibration integrals have been evaluated graphically for many cases by Davidson,[§] and Price^{||}. A_r is the

[†] 'Z. Physik,' vol. 41, p. 482 (1927).

[‡] Correspondingly $\bar{I}_{d^1\Pi_b \rightarrow 2p^1\pi} = \nu^4 [S_{11b}A_1B_1 + S_{12b}A_2B_2]^2$.

[§] 'Proc. Roy. Soc.,' A, vol. 135, p. 459 (1932).

^{||} 'Proc. Roy. Soc.,' A, vol. 136, p. 264 (1932).

electron-rotation factor having values for various transitions as given in Table I.

Table I.—Values for A_r .

	$d \rightarrow p\Pi$		
	A_0	A_1	A_2
$K \rightarrow K+1$	$\sqrt{(K+2)/(2K+1)}$	$\sqrt{3K(K+2)/(K+1)(2K+1)}$	$\sqrt{3K(K-1)/(K+1)(2K+1)}$
$K \rightarrow K$	-1	$\sqrt{3/K(K+1)}$	$\sqrt{3(K-1)(K+2)/K(K+1)}$
$K \rightarrow K-1$	$-\sqrt{(K-1)/(2K+1)}$	$\sqrt{3(K-1)(K+1)/K(2K+1)}$	$-\sqrt{3(K+1)(K+2)/K(2K+1)}$

	$d \rightarrow p\Sigma$	
	A_0	A_1
$K \rightarrow K+1$	$\sqrt{4(K+1)/(2K+1)}$	$\sqrt{3K(2K+1)}$
$K \rightarrow K$	0	$\sqrt{3}$
$K \rightarrow K-1$	$\sqrt{4K/(2K+1)}$	$-\sqrt{3(K+1)/(2K+1)}$

These values, using the Φ_{nLA} wave functions, are correct to within the order of the S_{nLA} 's discussed on p. 191.

It should be noted that all the Ψ_N 's forming the series for the "uncoupled" Ψ 's (equation (9)) have the same K, M values. Therefore Ψ is subject to the same K, M selection rules as is each Ψ_N , viz., $\Delta K = 0, \pm 1$, $\Delta M = 0, \pm 1$ (see Kronig's book, p. 70). The rule $\Delta \Lambda = 0, \pm 1$ breaks down since the Ψ_N 's in Ψ may involve differing Λ values, and thereby Λ loses its significance as a "true quantum number." Thus, with uncoupling, we may have the transition $3d^1\Delta_a \rightarrow 2p^1\Sigma$.

As was remarked at the beginning of this communication, Davidson will discuss the numerical application of this theory to the case of H_2 d^1 -states. There is good agreement between the simple theory and experimental data in many cases.

1.4. Note on a Method for including all the Vibration Effects.

We shall write the wave function Ψ_N (defined in equations (2) and (8)) in the form $\Psi_N = P_N(\rho) F_N(\xi, \eta, \zeta, \rho_e, \theta, \psi, s_e, s_n)$, and the vibration equation (5)

in the form $[G_N + W_N/B] P_N(\rho) = 0$. Then the variation equation (9) may be written :

$$\delta \left[\int_{r,s} \Sigma V_r^* \{ (G_s + W/B) \delta(r, s) - x_{rs} \} V_s \rho^2 d\rho \right] = 0. \quad (26)$$

Here, $V_r = \Sigma_{v_r} C_{N,r} P_{N,r}(\rho)$, summed over the complete system of orthogonal functions $P_{N,r}(\rho)$ associated with the electron-rotation function $F_{N,r}$. We have used $x_{rs} = \frac{1}{B} \int F_r^*(h' - h'_s) F_s d\tau_1$, where $d\tau_1$ refers to the complete configuration space excluding the ρ -variable. Equation (26) gives :

$$\Sigma_s [\{G_r + W/B\} \delta(r, s) - x_{rs}] V_s(\rho) = 0 \quad \text{for all } r\text{'s.}$$

This set of linear differential equations must be solved to determine the W 's and the V 's. For the pair $d^1\Pi_b$, $d^1\Delta_b$ the equations reduce to :

$$\begin{aligned} [G_1 + W/B] V_1 - 2\sqrt{(K-1)(K+2)} V_2 &= 0, \\ -2\sqrt{(K-1)(K+2)} V_1 + [G_2 + W/B] V_2 &= 0. \end{aligned}$$

If the potential energy curves $U_n(\rho)$ (see below equation (5)) for the two states coincide, these equations show that $W = W^*$, where $W^* \equiv W_r$ with K replaced by K^* , where $K^*(K^* + 1) = K(K + 1) \pm 2\sqrt{(K-1)(K+2)}$. Similar results hold for the case of the $d^1\{\Sigma, \Pi_a, \Delta_a\}$ states (as Dr. Davidson has shown the writer). These calculations indicate that the effect of considering only the $v_r = v_s$ elements in the determinants (22) and (23), is very nearly equivalent to including the effects of all the v_r, v_s elements.

II. THE STARK EFFECT IN H_2 .

2.1. Introduction.

In the Stark effect, strong electric fields acting on the source of a spectrum produce frequency displacements in some of the lines in the spectrum. A description of the observed effects in the spectrum of H_2 may be found in the writer's papers.[†] The Stark effect is very irregular for H_2 . Thus, the type of displacement may differ greatly between lines produced in nearly identical transitions, as, for instance, in lines associated with slightly differing initial vibration or rotation quantum numbers. In all cases where accurate measurement has been possible, the frequency displacements have been found to be very nearly proportional to the square of the field strength. We shall now develop formulæ for the Stark effect in H_2 .

[†] 'Proc. Roy. Soc.,' A, vol. 123, p. 103 (1929), and vol. 131, p. 146 (1931).

2.2. The Mathematical Problem.

We may insert the potential term due to the electric field, into equation (1) of Part I of this paper :

$$[H_0 + h' - FV_F - W] \Psi = 0, \quad (27)$$

F is the field strength ; V_F is the F -directed component of the electric vector V (as in equation (24)). For H_2 we have

$$V_F = eZ, \quad Z = [\{r_a \cos \theta_a + r_b \cos \theta_b\} \cos \theta + \{r_a \sin \theta_a \cos \phi_a + r_b \sin \theta_b \cos \phi_b\} \sin \theta], \quad (28)$$

where r_a , θ_a , ϕ_a are the spherical polar co-ordinates of electron "a" referred to the internuclear axis (origin at the centre). ϕ_a is measured from the plane containing the axis and running parallel to the direction of the field. θ is the azimuthal angle for the nuclear axis referred to the direction of the field.

Since the observed frequency changes caused by the field are small we may expect that the usual power series perturbation method will be legitimate here (at least in an asymptotic sense). Thus, the change $\Delta W'$ in the energy W' associated with the wave function Ψ' , is given by

$$\Delta W' = (Fe) \left[\int \Psi'^* Z \Psi' d\tau \right] + (Fe)^2 \left[\sum_{W'' \neq W'} \frac{\left| \int \Psi''^* Z \Psi' d\tau \right|^2}{(W' - W'')} \right] + O(F^3) \quad (29)$$

(there being no degeneracy which leads to an initial secular determinant). For H_2 the wave functions are such that $[\Psi'^* \Psi']$ has the property "g" for reflection of the electronic co-ordinates through the origin, while Z (equation (28)) clearly has the property "u." Therefore we find for H_2 that the field-perturbed energy and wave function have the forms :

$$[W' + \Delta W'] = W' + (Fe)^2 \left[\sum_{W''} \frac{|G''|^2}{(W' - W'')} \right] + O(F^4), \quad (30)$$

$$[\Psi' + \Delta \Psi'] = \Psi' \left[1 - \frac{1}{2} (Fe)^2 \left\{ \sum_{W''} \frac{|G''|^2}{(W' - W'')} \right\} \right] + O(F^4) + (Fe) \left[\sum_{W''} \frac{\Psi''^* G''}{W' - W''} \right] + O(F^2), \quad (31)$$

where

$$G'' = \int \Psi''^* Z \Psi' d\tau. \quad (32)$$

For our Ψ' , Ψ'' we must use the "uncoupled" wave functions, $\Psi = \sum_N C_N \Psi_N$.

of equation (9). (In many actual cases several of the larger C_N 's in the series for the "uncoupled" functions may be of comparable magnitude, hence it would usually be a poor approximation to use a simple form $\Psi = \Psi_N$ in calculating the G 's. However, when there appears to be little uncoupling, as in the case of the $3p^3$ -states, then the simple form $\Psi = \Psi_N$ may be used.)

Since Z in (32) is proportional to a component of the electric vector V used in the intensity formula (24), Ψ' and Ψ'' must be related as in transitions associated with polarisations parallel to F , in order to have non-zero G 's. Thus, $G'' = 0$ unless :—

- (i) $K'' - K' = 0, \pm 1$; $M'' - M' = 0$. (See paragraph above section 1.4.)
- (ii) Ψ'' and Ψ' have opposite reflection symmetries with respect to the origin.
- (iii) Ψ'' and Ψ' have the same co-ordinate interchange symmetries.

The $2K + 1$ -fold degeneracy associated with the M 's for a given K is largely removed by the field. Separated "components" of the original level are formed thereby. This $(2K + 1)$ -fold degeneracy does not, of course, produce a $(2K + 1)$ -rowed "first order" secular determinant since the wave functions corresponding to different M 's in the degenerate set give zero G'' values (by (i) above).

2.3. General Theoretical Properties of the Stark Effect in H_2 .

(1) The levels associated with $+M$ and $-M$ are equally affected by the field. This property corresponds to the physical picture in which the two senses of rotation of the molecule about the field direction are equivalent as regards the effect of the field. Mathematically this $\pm M$ degeneracy is associated with the fact that integrations over the rotation angles in G'' give factors involving M only as M^2 (see section 2.4.)

(2) The interchange symmetries are unaffected by the field. This property follows from the fact that the perturbation term, FV_F in (27), is symmetrical for interchanges of the co-ordinates of like particles, and therefore cannot affect the symmetry properties of the equation.

(3) For the normally permitted transitions, with and without the field, we have the selection rules $\Delta K = 0, \pm 1$, and $\Delta M = 0$ for polarisations parallel to the field, $\Delta M = \pm 1$ for polarisations perpendicular to the field. The various components associated with the different M 's for a given K , unite at zero field strength. Using $|M| \leq K$ and (1) above, we may form Table II.

Table II.—Maximum Possible Numbers of Separate Components of Lines produced in Normal Types of Transitions.

	$\Delta M = 0 (\equiv).$	$\Delta M = \pm 1 (\equiv \perp).$
$K \rightarrow K + 1$	$K + 1$ components	$2K + 1$ components
$K \rightarrow K$	$K + 1$ "	$2K$ "
$K \rightarrow K - 1$	K "	$2K - 1$ "

Some of these components may have zero intensity or may coincide with other components for small field strengths.

(4) In addition to the "normal" transitions mentioned above, there are normally "forbidden" transitions brought into existence by the field. These "forbidden" transitions are associated with departures from the rule, "Combining states must have opposite reflection symmetries,"† and the rule $|\Delta K| \leq 1$. The resulting spectral lines will have frequencies not corresponding to any in the normal spectrum (except in accidental cases). These new lines are associated with the presence of the Ψ'' functions (and also the $O(F^2)$ term) in the perturbed wave function (31). By condition (i), p. 198, we see that Ψ'' may involve rotation functions with $K'' = K' + 1$. Therefore, as consideration of the appropriate intensity formula shows, the level concerned may combine with another level represented by a rotation number equal to $K'' + 1$. Thus we may have the transition $\Delta K' = 2$. Again, (ii), p. 198, requires that the perturbed wave function involve Ψ''' 's having reflection symmetries opposite to those of the unperturbed function, and thereby these properties lose their significance in the presence of the field. This feature is a general characteristic of the Stark effect for all atomic and molecular systems. As an example we may note that the normally forbidden transition $3p\ ^3\Pi \rightarrow 2p\ ^3\Pi$ can occur when there is a field present.

2.4. Numerical Formulæ for the Stark Effect in H_2 .

When there is sufficient experimental knowledge of the spectrum of H_2 in the absence of the field, the uncoupling theory given in Part I of this paper may be used to determine the C_N 's in $\Psi = \sum_N C_N \Psi_N$ (equation (9)). Also, when experimental data establish relationships between transitions from the levels W' and W'' in the normal spectrum, the quantity $(W' - W'')$ may be

† Kronig, *loc. cit.*, p. 72, and Mulliken, 'Rev. Mod. Phys.,' vol. 2, p. 153 (1930).

determined accurately. Under these circumstances we could calculate the Stark effect displacements

$$\Delta W' = (Fe)^2 \left[\sum'_{W'' \neq W'} |G''|^2 / (W' - W'') \right], \quad (33)$$

providing that we could calculate the integral, $Z_{N''N'} = \int \Psi_{N''}^* Z \Psi_{N'} d\tau$, which occurs in

$$G'' = \int \Psi''^* \Psi'' Z d\tau = \sum_{N', N''} \left\{ \int C_{N'}^* C_{N''} \Psi_{N''}^* Z \Psi_{N'} d\tau \right\} = \sum_{N', N''} C_{N'}^* C_{N''} Z_{N''N'}. \quad (34)$$

We shall now consider the integral $Z_{N''N'}$. (To obtain the positive signs given for the ω 's in Part I, we must use suitable normalisation phases. Thus $\Psi_{N'}$ must involve a phase factor $\alpha_{\Lambda'}$ (defined below) when we use $\Theta_{K, M, \Lambda}$ in the form given in equation (4), and when ϕ_a is measured as described below equation (28). The α 's are determined by the relations $\alpha_0 = 1$ and $\alpha_{\Lambda+1} = \pm \alpha_{\Lambda}$ according as $M \gtrless \Lambda + \frac{1}{2}$. With this convention we obtain the following results.)

Using the symmetry properties of the Ψ_N 's and of Z , we have

$$Z_{N''N'} = (x_{\Lambda'', \Lambda'}) (r_0) (R_{KM\Lambda}) (E_{q, \Lambda, K, v}). \quad (35)$$

Here

$$x_{\Lambda'', \Lambda'} = \sqrt{2},$$

if only one of the numbers Λ'' or Λ' is zero, otherwise, $x_{\Lambda'', \Lambda'} = 1$

$$r_0 = \hbar^2 / 4\pi^2 m e^2.$$

The factor $(R_{KM\Lambda})$ arises in the integrations over the rotation angles. Considering the phases and using the methods of Rademacher and Reiche (*loc. cit.*) we have in Table III the values of $(R_{KM\Lambda})$ for the non-zero $Z_{N''N'}$'s.

The factors in the second column are to be multiplied into the other factors in a horizontal line, as indicated by the \times sign, to give the actual $(R_{KM\Lambda})$ values.

The factor $(E_{q\Lambda Kv})$ comes from integrations over the electronic and the vibration co-ordinates thus :

$$(E_{q\Lambda Kv}) = 2 \int P_{q'\Lambda''K''v''}^* (\rho) \Phi_{q'\Lambda''}^* \left\{ \frac{r_a}{r_0} \right\} \{ \cos \theta_a + \sin \theta_a \cos \phi_a \} \\ \times \Phi_{q'\Lambda'} P_{q'\Lambda'K'v} d r_{\xi\eta\zeta} \rho^2 d\rho. \quad (36)$$

One should integrate first over the electron co-ordinates to obtain a function of ρ only, in the integrand ; then one should perform the ρ integration. How-

$$R_{KMA} = \alpha_A \alpha_{A'} \int \Theta^*_{K''M''A'} \{ \cos \theta \delta(\Lambda'', \Lambda') + \sin \theta \delta(\Lambda'', \Lambda' \pm 1) \} \Theta_{K'M'A} \sin \theta d\theta d\psi.$$

Values of (R_{KMA}) (see below for explanation).

Table III.

$(K'' - K')$ ↓	$(\Lambda'' - \Lambda') \longrightarrow$	+1	0	-1
+1	$\sqrt{\frac{Q(K'', M)}{K''(2K''+1)}} \times$	$\left(\sqrt{\frac{O(K'', \Lambda'+1)}{K''(2K''-1)}} \right)$	$\left(\sqrt{\frac{Q(K'', \Lambda')}{K''(2K''-1)}} \right)$	$\left(\sqrt{\frac{O(K'', -\Lambda'+1)}{K''(2K''-1)}} \right)$
0	$\sqrt{\frac{Q(M, O)}{K'(K'+1)}} \times$	$\left(\sqrt{\frac{P(K', \Lambda')}{K'(K'+1)}} \right)$	$\left(\sqrt{\frac{Q(\Lambda', O)}{K'(K'+1)}} \right)$	$\left(\sqrt{\frac{P(K', -\Lambda')}{K'(K'+1)}} \right)$
-1	$\sqrt{\frac{Q(K', M)}{K'(2K'-1)}} \times$	$\left(\sqrt{\frac{O(K', -\Lambda')}{K'(2K'+1)}} \right)$	$\left(\sqrt{\frac{Q(K', \Lambda')}{K'(2K'+1)}} \right)$	$\left(\sqrt{\frac{O(K', \Lambda')}{K'(2K'+1)}} \right)$

In this table, $(-)$, $(+)$, etc., refer to: $\frac{M}{M} \geq 0$.

$$Q(x, y) = x^2 - y^2; \quad O(x, y) = (x+y)(x+y-1); \quad P(x, y) = (x-y)(x+y+1).$$

ever, if we use the Φ_{nLA} functions as in Part I and if we neglect quantities like S^2_{nLA} , then we get:

$$(E_{q\Delta K v}) = (D_{nL\Delta K v})(E_{nLA}),$$

where

$$(D_{nL\Delta K v}) = \int P^*_{q''\Lambda''K''v''}(\rho) P_{q'\Lambda'K'v'}(\rho) \rho^2 d\rho, \quad (37)$$

and

$$(E_{nLA}) = \int \langle n''L''\Lambda'' | a \rangle^* \{ \cos \theta_a + \sin \theta_a \cos \phi_a \} \frac{r_a}{r_0} \langle n'L'\Lambda' | a \rangle d\tau_a \quad (38)$$

The type of vibration integral (37) has been considered graphically for many cases by Davidson and by Price (*loc. cit.*). When the nuclear potential functions are nearly the same for the two states N' , N'' , as is usually the case for excited levels, then $(D_{nL\Delta K v}) \doteq \delta(v'', v')$.

For the integral (38), we use:

$$(E_{nLA}) = (X_{nL})(Y_{LA}) \quad \text{with} \quad (X_{nL}) = \int_0^\infty R_{n''L''}(r) R_{n'L'}(r) r^2 dr, \quad (39)$$

where

$$R_{nL} = \sqrt{\left\{ \frac{(n-L-1)!}{2n[(n+L)!]^3} \left(\frac{2}{n}\right)^3 \right\}} \left(\frac{2r}{n}\right) L e^{-r/n} \left[L^{\frac{2L+1}{n+L}} \left(\frac{2r}{n}\right) \right] (2L+1)^{-\frac{1}{2}}.$$

The values for Y_{LA} are listed in Table IV.

Table IV.

$(L''-L') \downarrow$	$(A''-A') \rightarrow$	+1	0	-1
+1		$\frac{1}{2} \sqrt{(L''+A')(L''+A')}$	$\sqrt{(L'')^2-(A')^2}$	$-\frac{1}{2} \sqrt{(L''-A')(L''-A')}$
-1		$-\frac{1}{2} \sqrt{(L'-A')(L'-A')}$	$\sqrt{(L')^2-(A')^2}$	$\frac{1}{2} \sqrt{(L'+A')(L'+A')}$

The integral (X_{nL}) in (39) may be calculated for any case, but since the $1/(W' - W'')$ values which occur in the expression (33) will be large for N' and N'' corresponding to the same principal quantum number n (orbital "degeneracy"), we may expect that only the $n'' = n'$ elements need be considered. For this case we have

$$(X_{nL}) = -\frac{2}{3} n' \sqrt{[(n')^2 - (L_g)^2] / [2L' + 1][2L'' + 1]}$$

for $n' = n''$ and $L'' = L' \pm 1$ (40)

L_g being the larger of L' and L'' .

Intensity Formulae for the Stark Effect Components.—To calculate the intensities we must use the perturbed functions (31) in the intensity formula similar to equation (24). For polarisations parallel to the field we must use the F-directed component of the electric vector, and for polarisations perpendicular to the field we must use a component of the electric vector perpendicular to the field. These components will replace the V in equation (24). Thus, for the normally permitted transitions, say, between Ψ' and Ψ'' , we have the intensities given below (neglecting terms in F^2):

$I_p = 4\nu^4 e^2 |G''_p|^2$ for "parallel components" (where $G''_p = G''$ has the form (34));

$I_s = 4\nu^4 e^2 |G''_s|^2$ for "perpendicular components" (where the G''_s -values have the same form as the G'' 's except that the "M-factors" given in the second column of the table for the (R_{KMA}) -values, must be changed to the following:—

$(K'' - K')$	$(M'' - M') = \pm 1.$
+1	$\frac{1}{2} \sqrt{\frac{O(K'', \pm M' + 1)}{K''(2K'' + 1)}}$
0	$\frac{1}{2} \sqrt{\frac{P(K', \pm M')}{K'(K' + 1)}}$
-1	$\frac{1}{2} \sqrt{\frac{O(K', \mp M')}{K'(2K' - 1)}}$

The extension of the formulæ to the case of the "forbidden" transitions presents no difficulties since G''_p and G''_s forms again appear when one considers the contributions of the intensity integral from the functions equivalent to Ψ'' in the perturbed wave function (31). It will be instructive to consider an example of "forbidden" transitions. Let us take the

$$\{3p\ ^3\Pi_a, K=2(\equiv J=2\frac{1}{2}), |M|=2, v=0\}$$

level considered by Richardson and Das.[†] This level (or, more correctly, this spin complex) is associated with positive symmetry for nuclear interchanges. Calling the unperturbed wave function Ψ' , as in equation (31), we have Ψ'' functions in the perturbed part which correspond to the following $3d^3$ -states described by Richardson and Davidson[‡]:

$$3d^3 \{[\Sigma J=2\frac{1}{2}; \Pi_a J=2\frac{1}{2}; \Pi_b J=3\frac{1}{2}; \Delta_a J=2\frac{1}{2}; \Delta_b J=3\frac{1}{2}] |M|=2, v=0\}.$$

The level $\{3d\ ^3\Pi_b, J=1\frac{1}{2}(\equiv K=1)\}$ is ruled out for our purposes since we must have $K \geq 2 = |M|$. Thus, the presence of these uncoupled Ψ'' functions will permit transitions from the perturbed $\{3p\ ^3\Pi_a, K=2, |M|=2, v=0\}$ level to the final levels possible for the above $3d^3$ states, as given in Richardson and Davidson's "fig. 1" (*loc. cit.*). The intensity for such a transition will be given by

$$I = 4\nu^4 (Fe)^2 \left| \frac{G''}{W' - W''} \right|^2 |P|^2,$$

where P has the form G''_p or G''_s with the Ψ'' (in the G'' -form) replaced by the wave function for the final state, and the Ψ' replaced by the wave function for the $3d^3$ -state under consideration.

Unfortunately, as yet there is not sufficient knowledge of the spectrum of H_2 (in the absence of the field), to permit the numerical application of the theory of the Stark effect. Thus, for the triplets there is some confusion about the d^3 -states and the uncoupling theory cannot be applied there yet. For the singlets, the $3p^1$ and $4p^1$ levels have not been analysed spectroscopically yet, and there are certain "extra levels" which require investigation. However, there are some general properties of the observed Stark effect patterns in H_2 which are predicted by the theory (see the writer's paper (*loc. cit.*)).

[†] 'Proc. Roy. Soc.,' A, vol. 122, p. 697 (1929).

[‡] 'Proc. Roy. Soc.,' A, vol. 131, p. 673 (1931).

§ That is, compatible with $M \leq K$ for the final state.

It is hoped that the present communication will encourage investigation of the spectroscopic and the theoretical properties of the excited levels in H_2 .

The writer expresses his indebtedness to Dr. P. M. Davidson for helpful discussion in the course of the investigation, and acknowledges, with thanks, supervision by Professor R. H. Fowler. The investigation was carried out by means of grants from the Royal Commission of the 1851 Exhibition, and Emmanuel College.

Summary.

The wave equation for diatomic molecules is expressed in a form suitable for light molecules. A theory for successive approximations to give the energy values is developed. Only few-rowed determinants appear in the calculations. Expressions for calculating the gyroscopic effect called "uncoupling" are developed for the d^1 -states in H_2 ; the numerical application of these expressions in comparison with experimental material is considered by Davidson in a separate communication. A theory for the Stark effect in H_2 is developed on the basis of the preceding theory; general formulæ are given for calculating the intensities and the displacements of Stark effect components. Complete numerical calculations for the Stark effect, cannot be made yet, due to lack of knowledge of the normal spectrum.

The Polarity of Thunderclouds.

By E. C. HALLIDAY, Ph.D., Lecturer in Physics, Witwatersrand University.

(Communicated by C. T. R. Wilson, F.R.S.—Received June 10, 1932.)

[PLATES 5–10.]

1. *Introduction.*

The mechanism of the maintenance of the negative charge upon the surface of the earth has long been sought. C. T. R. Wilson has made the suggestion that the activity of thunderstorms of positive polarity—positively charged above and negatively charged below—will serve to separate positive and negative charges, by carrying negative charges to the earth and positive charges to the upper atmosphere. Experiments carried out by him during and subsequent to 1916 indicated the presence of clouds predominantly of positive polarity,* and similar experiments carried out by Schonland and Craib, in South Africa, gave similar conclusions.†

Appleton, Watt and Herd‡ made observations on the form of atmospherics, and came to the conclusion that the thunderclouds which were the seat of the disturbances producing atmospherics were of positive polarity.

Again, Wormell§ from the results of a series of experiments on the interchange of electricity between stormclouds and the earth, considered that thunderclouds were predominantly of positive polarity.

A different view as to the distribution of thundercloud-charges has been put forward by G. C. Simpson.||

In this paper are described experiments carried out in Johannesburg, in South Africa, in which the polarity of thunderclouds was investigated by making correlations between the nature of lightning flashes and the sign of the sudden field changes which they produced.

2. *Site and Apparatus.*

The Physics Department of the University of the Witwatersrand occupies the top story of a two-storied building standing upon a ridge which runs east

* 'Proc. Roy. Soc.,' A, vol. 92, p. 555 (1916); 'Phil. Trans.,' A, vol. 221, p. 73 (1921).

† 'Proc. Roy. Soc.,' A, vol. 114, p. 229 (1927); 'Proc. Roy. Soc.,' A, vol. 118, p. 233 (1928).

‡ 'Proc. Roy. Soc.,' A, vol. 111, p. 654 (1926).

§ 'Proc. Roy. Soc.,' A, vol. 115, p. 443 (1927).

|| 'Phil. Trans.,' A, vol. 209, p. 379 (1909); 'Proc. Roy. Soc.,' A, vol. 114, p. 376 (1927).

and west through the City of Johannesburg about 2 miles north of the line of mine dumps which marks the position of the gold reef. This gives an uninterrupted view to the south, south-west and south-east, but to the west the view is partially obscured by the university buildings and to the east the rising ridge obscures the horizon.

The apparatus was housed in a small laboratory in the south-east corner of the building with two windows facing south. An exposed test-plate such as that used by Wilson and Schonland was not very practicable, as the room was so much above the ground. Instead, an insulated copper sphere was suspended from a cable stretched from the roof of the building to a standard in a field to the south of it. The ball hung from a point along the cable so that it was 4 metres from the wall of the building and 9 metres from the ground. It was 33 cm. in diameter and was suspended from a sulphur-ebonite insulator protected from the weather. A wire soldered to the ball passed into the laboratory through a hole in one of the window panes. The ball served to record the changes of field produced by flashes near the station, but for more distant storms an aerial was used. The latter stretched from a point on the wall of the building 7 metres high to a standard in the field to the south of the building, and was 18 metres long. The insulators were also of ebonite, with sulphur guard-rings protected from rain, and the "lead-in" passed through another pane in the window.

The recording apparatus consisted of a capillary electrometer of the type originally used by Wilson* and a device whereby the movements of the capillary were registered on a moving sensitive plate. It was the same apparatus as had been used by Dr. Schonland in Somerset East and was lent by him for the research.

The chief aim of the investigation was to obtain information about the sign of the change of field produced by different types of flashes. To do this it was necessary to have a means of watching the flashes and at the same time making a mark on the record to indicate which field change took place at the same time as the flash whose nature was observed. The method used enabled the "flash observer" to concentrate nearly all his attention upon watching the flashes, as the only other operations which he had to perform were to press a key which marked the record with a white line, and note the nature of the discharges in sequence in the log book. To accomplish this the following device was used. The lamp used to illuminate the electrometer was a 50-c.p. 12-volt

* 'Phil. Trans.,' A, vol. 221, p. 76 (1921).

motor car head lamp. This lamp was supplied with 16 volts through a suitable resistance. The "make and break" of a relay was connected across the resistance and the relay was connected to the 16 volt supply through a long coil of flex wire wound on a drum and having its free ends connected to a key. When a storm was approaching this wire was run out to any window which afforded a clear view of the flashes. When the operator of the recorder started an exposure he rang an electric bell. The "flash observer" then watched for flashes, and whenever he saw one, pressed the key and also made a note of the nature of the flash. For each new plate in the recorder a new series of numbering was commenced, and so after the storm was over and the plates were developed, it was easy to tell which mark on the plate corresponded to each flash noted in the observer's log book. Brightening the light was found to be better than switching out the light, as it did not cause any discontinuity in the record and so the progress of recovery of the field could be seen, and any flash occurring immediately after the first one was not obliterated. This method of making correlations is superior to any system of timing, for there is no doubt about the identity of the field change indicated. When flashes are taking place at the rate of 15 to the minute, as was sometimes the case, or if there were any mistake about the exact second at which a record was started, it would be quite easy for a flash to be correlated with the wrong field change, but with a system of marking, the flash is identified quite unmistakably; the mark on the record appearing immediately after the field change. In addition it was possible to detect flashes which produced no field change, for the average time lag of the observer being known, if there was no field change at a distance ahead of the observer's mark equal to, or slightly greater than this time lag, then it was clear that the flash observed could not have caused any change of field. In practice it was found that it was a simple matter to watch the sky for flashes, press the key as soon as one was seen and then write down "earth" or "cloud" on the log sheet without looking at the paper, and sometimes add notes such as "very large flash" or "flash very inclined." With a stop-watch the time interval between the flash and the thunder following it was measured as often as possible and noted in the log book, but this could not be done with every flash observed, especially when flashes came in quick succession.

When storms took place at night, photographic records of the nature of the flashes were made, but this also could not be done for every flash as the film had to be changed for each exposure so the number of correlations made photographically is not large.

3. *A General Account of the Conditions under which Observations were made.*

During the period October, 1928, to December, 1930, 38 storms were observed. Most of them were first seen in a S.W. direction moving in a general N.E. direction which usually brought them quite near the station and occasionally carried them right over it. A large proportion of them occurred during the daylight hours of 1 to 5 p.m. so that the observations have been affected in two ways :—

- (1) The storms were very seldom observed at distances greater than 14 km. In the daylight it is difficult to pick up the lightning from a storm at a distance visually. Once the lightning has been seen it is easy to follow the progress of the storm to greater distances, but it is difficult to observe the beginning of it at such distances. Thunder is heard only at about 15 km. and nearer, and with the sounds of traffic all around it is difficult to distinguish the thunder from other sounds. By the time the thunder is heard quite clearly, the distance of the storm is about 7 km. Thus unless a very close and constant watch is kept, daylight storms are very unlikely to be observed until they are nearer than 14 km., and with the work of the department to be carried out a close watch could not always be kept.
- (2) The observation of flashes in a daylight storm is always difficult. The intensity of the flashes relative to the general illumination is less than at night and as a result only powerful flashes are seen. The result is that flashes to the ground are almost always correctly observed while there is an element of uncertainty about flashes in the clouds where the intensity is weakened by the screening effect of the cloud. This is important in the case of double flashes of lightning, as is discussed in a later paragraph. There is also a greater possibility by day than by night of an observer seeing the flicker produced by a flash to the ground, but failing to see the actual flash owing to its being on the edge of his field of vision, and reporting it as a cloud flash. It is not likely that such a happening was of frequent occurrence as lightning flashes usually take place in a fairly restricted part of the sky, but it must be recognised that the observation of flashes is not entirely free from error and allowance must be made for such possible error. Further reference is made to this point later.

Wilson, Schonland and Wormell refer to the possible difficulty of the interfering effects of two or more storms in action at the same time. The extent to

which this difficulty was sometimes met with may be judged by the fact that on certain occasions an observer on the roof of the building, from which an uninterrupted view could be obtained, reported that flashes could be seen in a complete semicircle stretching round the horizon from west to east. Some of these flashes were near and some were far and it was a matter of some difficulty to tell if the flash seen were really responsible for the field change produced.

The timing of the interval between the flash and the thunder is an important part of the observations, as without it no accurate estimate of the distance of the storm can be made, and a correct interpretation of the records depends upon a knowledge of the distance of the storm. During some of the storms the flashes followed so fast after each other that timing of the thunder interval was impossible and valuable information was lost.

Another troublesome phenomenon was the presence of large fields when the active centre of the storm was still far away. It sometimes happened that a storm was preceded by a charged cloud or even the remains of a previous storm, and the field produced by this charged mass overhead was sufficient to cause point discharge from the aerial, causing the capillary bubble to move with a steady velocity and making all attempts to record field changes unsuccessful. This sometimes happened when the flashes were taking place at such a distance that the ball was not sensitive enough to record the field changes produced by them. A method was devised which enabled observations to be made under such circumstances. The electrometer was shorted by a dial resistance box with a maximum resistance of 1 megohm. Usually this resistance was kept at the infinity mark, but when a point discharge took place it was given such a value that the current was by-passed through the resistance and the electrometer was not disturbed. This was possible because there is a minimum current through the electrometer below which no movement takes place. The sudden large momentary current produced by a flash of lightning passed more readily through the electrometer, as the resistance box possessed a slight self-inductance. Thus the field change was registered, though the magnitude of the displacement of the mercury was reduced. The resistance box had the disadvantage of seriously affecting the recovery curves, but it enabled the sign of the field change to be observed and this was the main object of the research. A practical test of the apparatus was made, in which the capillary electrometer was subjected to an alternating e.m.f. with a frequency of 5 per second in one case and 10 per second in another, while the shunting resistance was connected across the electrometer terminals, and a photographic record

made of the capillary movements showed that they faithfully followed the reversals of sign of the potential applied.

4. *Scope of the Investigation.*

C. T. R. Wilson's experiments on the fields near thunderstorms led him to the theory that thunderclouds are essentially bi-polar, the positive charge being above the negative charge.* The simplest type of such a cloud would have the two charges in a vertical line, but in the discussion of the results, Wilson considers the possibility of the two charges being displaced laterally and the effect which such a disposition of charge would have on the field near the cloud.

Schonland summarised the discussion of Wilson and placed it in tabular form.* There are four different types of flash to be expected from a thundercloud of positive polarity, each of which will produce a field change at a station in the neighbourhood of the cloud :—

- (1) A flash from the bottom of the cloud to the ground will always produce a positive field change.
- (2) A flash from the top of the cloud to the ground will always produce a negative field change.
- (3) A flash from the top of the cloud into the upper atmosphere, dissipating positive electricity into the upper conducting layer will always produce a negative field change.
- (4) A flash from the positive top of the cloud to the negative bottom of the cloud will produce a field change which changes from negative to positive as the storm approaches the station.

Schonland did not think that flashes (2) and (3) were likely to take place as conditions did not favour them, and in the case of flash (3) practically no evidence exists of such a flash ever having been seen. There remain flashes (1) and (4), and a study of the field changes due to these two types of flash led Schonland to infer that the majority of the thunderstorms studied were of positive polarity. In 12 storms occurring at a distance greater than 15 km. 523 flashes were observed in the clouds, of which 517 produced negative changes and 6 produced positive changes of field.

Out of 54 positive field changes due to distant flashes, 48 were due to flashes to the ground and 6 were due to flashes in the cloud.

* 'Proc. Roy. Soc.,' vol. 114, p. 233 (1927).

The work described in this paper has been carried out to add to the general body of information on the subject of the polarity of South African thunderstorms. Information has also been obtained on other points of interest discussed by Schonland such as (a) the reversal of the sign of the majority of the field changes with distance ; (b) the sign of the field changes from overhead storms ; (c) the steady fields due to near and distant storms. These data will be discussed in a later paper together with observations made subsequent to December, 1930.

5. General Record of Storms Observed.

Below is given a general record of the storms observed. The first column gives the reference letter of the storm. Where figure suffixes are given it is because during the observation of the storm its distance from the station became less and the nature of the field changes altered. The interpretation of the records is facilitated by dividing the storm up so that the effect of distance is shown. The second and third columns give the date and time of day when the observations were made. Column 4 gives the distance of the storm from the station. Columns 5 to 8 give a record of the field changes which were correlated with observation of the flashes and show : (5) cloud flashes which produced positive changes ; (6) cloud flashes which produced negative changes ; (7) flashes to ground which produced positive changes ; and (8) flashes to ground which produced negative changes. The last two columns show the total number of field changes which have been recorded, including those correlated with observations of the flashes and those where no correlation was made. In column 4 the symbol (OH) means that the storm was directly above the station. A flash in the clouds above the station will usually produce thunder after an interval of several seconds and will therefore give the impression that it is 1 or 2 km. distant. Therefore some of the storms which are marked 3 km. or 4 km. were probably nearer to the station in a horizontal line, but it was decided to mark as overhead storms only those storms where the lightning flashes were actually seen above the station.

It will be seen that a total number of 3,015 field changes were recorded, of which 560 were correlated with observations of the flashes. This gives a ratio of 5.4/1 for the number of flashes taking place to the number of flashes seen, so that out of every six flashes in the storm at least one was seen. This is a fair sampling of the flashes taking place, and it is not likely that a more complete and thorough observation of the storms, if that had been possible, would change the nature of the results.

Table I.—General Record of Storms Observed.

Storm.	Date.	Time.	Distance.	Correlation with outside observation.				Total.	
				Cloud flash.		Ground flash.			
				ΔF+	ΔF−	ΔF+	ΔF−	ΔF+	ΔF−
A	14.10.28	22.40/23.30	km. 30	0	18	2	0	5	40
B	15.10.28	22.30/23.10	30	0	7	1	0	4	14
C ₁	22.2.29	20.00/20.40	10	0	11	3	0	19	35
C ₂			4	16	3	7	1	35	5
D	18.3.29	16.40/17.10	3	0	0	0	0	12	1
E	20.3.29	21.20/22.00	12	1	51	14	1	15	52
F	27.3.29	21.20/22.20	10	0	0	0	0	23	45
G	28.3.29	16.50/17.05	4	0	0	0	0	2	18
H	30.4.29	22.00/23.00	11	1	20	5	0	15	32
I	6.6.29	21.15/21.50	4	0	0	0	0	6	14
J ₁	9.10.29	19.30/20.00	(OH)	3	1	5	0	8	1
J ₂		20.00/22.00	8	0	1	1	0	20	18
K	24.10.29	10.30/ ?	1.5	0	0	0	0	13	10
L ₁	27.10.29	22.20/22.52	2	0	0	0	0	7	7
L ₂		23.00/23.45	6	0	0	0	0	0	48
M	28.10.29	13.15/14.00	12	1	4	17	1	18	5
N	8.11.29	evening	8	0	0	0	0	0	15
O ₁	6.12.29	14.45/15.00	9	0	3	17	0	17	3
O ₂		15.00/15.30	2	0	0	0	0	27	5
P	29.12.29	18.06/18.30	7	0	0	0	0	0	11
Q	30.12.29	13.07/13.09	?	0	0	0	0	4	14
R ₁	5.1.30	16.43/17.01	8	3	1	4	0	28	73
R ₂		17.03/17.14	6	1	1	1	0	29	24
S	9.1.30	15.56/17.01	3	5	2	2	1	60	13
T ₁	11.1.30	17.24/18.11	8	0	0	5	0	8	35
T ₂		18.27/19.11	4	8	0	0	0	55	0
T ₃		21.00/ ?	20	0	4	1	0	1	11
U ₁	24.1.30	13.18/13.37	6	0	0	22	0	32	27
U ₂		13.38/14.03	4	0	0	18	0	74	26
V	27.1.30	14.11/14.21	8	0	0	2	0	5	35
W	28.1.30	14.27/14.53	4	2	0	13	1	65	20
X	29.1.30	16.07/16.31	4	0	0	0	0	37	30
Y	30.1.30	17.01/ ?	11	0	0	11	1	23	78
AA ₁	21.10.30	16.17/16.42	8	0	1	3	0	7	11
AA ₂		16.52/17.40	4	3	6	0	0	10	14
AB	22.10.30	13.28/14.27	13	0	0	13	0	65	112
AC	1.11.30	20.55/21.10	5	3	1	1	0	6	1
AD ₁	6.11.30	14.38/15.16	6	6	0	0	0	136	15
AD ₂		15.32/16.50	(OH)	10	1	0	0	41	4
AD ₃		18.08/18.25	2	2	1	0	0	2	1
AD ₄		18.33/18.37	7	0	4	2	0	8	17
AE ₁	14.11.30	17.25/18.04	3	3	0	0	0	22	0
AE ₂		18.32/18.37	8	0	0	0	0	1	4
AF ₁	19.11.30	12.47/13.16	11	0	0	8	1	20	20
AF ₂		13.33/14.18	4	2	0	3	0	20	2
AG	20.11.30	17.00/17.41	5	11	0	10	0	41	4
AH	27.11.30	14.55/15.20	4	3	0	2	0	20	1
AI	4.12.30	15.22/16.20	17	2	1	3	0	24	106
AJ ₁	5.12.30	11.54/12.40	10	0	0	0	0	3	20
AJ ₂		12.41/13.13	4	5	1	1	0	17	9
AK	6.12.30	17.26/19.14	12	3	14	4	0	45	191
AL	7.12.30	17.58/20.09	7	6	3	17	2	63	37
AM ₁	8.12.30	17.52/18.43	15	1	4	39	7	83	262
AM ₂		18.48/19.34	4	3	1	6	0	63	15
AM ₃		19.36/19.54	6	2	6	4	0	17	20
Total flashes observed = 3,105.				106	171	267	16	1,392	1,623

In the table a line has been drawn below storm "O₂." All the storms above this line were observed by visual recording of the field changes, using a microscope focussed on a capillary electrometer. All the storms below the line were recorded on the photographic recorder.

6. *Correlation of the Appearance of the Flashes with the Sudden Field Changes which accompanied them.*

(a) *Distant Flashes within the Cloud.*—In Table II all the flashes which were observed to take place in the cloud without any discharge to earth have been collected. The first column gives the reference letter of the storm as shown in the general table, the second gives the distance in kilometres and the last two give the number of field changes of either sign accompanying the flashes.

Table II.—Field Changes due to Distant Flashes in the Clouds.

Storm.	Distance.	Field changes.	
		Positive.	Negative.
	km.		
A	30	0	18
B	30	0	7
C ₁	10	0	11
E	12	1	51
H	11	1	20
O ₁	9	0	3
M	12	1	4
T ₂	20	0	4
AA ₁	17	2	1
AK	12	3	14
AM ₁	10	1	4
		9	137

The total is 137 negative and 9 positive field changes accompanying flashes of lightning in the clouds. The ratio of 15/1 in favour of a flash in the clouds producing a negative change of field is good evidence that the storms were of positive polarity. The distance at which cloud flashes begin to produce positive field changes has been estimated by Schonland as about 12 km., though the presence of inclined flashes produces a zone extending from the 7 km. mark to the 12 km. mark in which field changes of either sign are to be expected. At Johannesburg, 6,000 feet above sea level, the clouds are nearer to the earth than at lower altitudes, and the outer limit of the intermediate zone is smaller than 12 km. It was decided therefore to include

storms which were as near as 9 km. and it will be seen from the table that 7 out of the 9 positive field changes were produced by storms 12 and 17 km. away and only 2 were produced by storms 10 and 11 km. away.

(b) *Flashes to the Ground*.—Table III gives a summary of the field changes which accompanied all flashes to the ground at any distance from the station. The columns give the index letter of the storm, the number of positive field changes associated with flashes to the ground, and the number of negative field changes due to similar flashes.

Table III.—Field Changes due to Flashes to Ground at any Distance.

Storm	Field Changes.	
	Positive.	Negative.
A	2	0
B	1	0
C	10	1
E	14	1
H	5	0
J	6	0
M	17	1
O	17	0
R	5	0
S	2	1
T	6	0
U	40	0
V	2	0
W	13	1
Y	11	1
AA	3	0
AB	13	0
AC	1	0
AD	2	0
AF	11	1
AG	10	0
AH	2	0
AI	3	0
AJ	1	0
AK	4	0
AL	17	2
AM	49	7
	267	16

This table shows that out of 283 flashes to ground observed, 267 produced positive field changes and 16 produced negative field changes, giving a ratio of 16/1 in favour of the charge of electricity brought to earth by a flash of lightning being a negative one. This is a very solid piece of evidence that the lower portions of a thundercloud carry a negative charge. It must be noted that out of the 16 negative field changes 7 are due to one storm (AM) and it is

possible that this storm may have been a special case. It is one of the storms in which the activity was very widespread, notes being made that flashes had been seen from several storm centres, and it may be that the negative changes were caused by flashes which were obscured from view by the college buildings but which produced larger field changes than the flashes which took place in another centre at the same time and were seen. Such a coincidence could not be expected to take place a very large number of times, but when all the lightning flashes could not be seen it is quite possible that such an error of observation might take place, especially as the storm produced a high frequency of flashes, 460 being recorded in 2 hours. The storm gave a preponderance of positive over negative field changes but produced an unusual number of negative field changes; thus the evidence as to its being of positive polarity is weak. If it is eliminated from the table the ratio becomes positive/negative 24/1.

(c) *Flashes in Very Near Clouds.*—Table IV gives the collected observations on cloud flashes in all storms which were nearer than 6 km. to the station, and some which were definitely overhead.

The columns contain (1) the reference letter of the storm as given in the general table, (2) its distance from the station, (3) the number of positive field changes due to flashes in the clouds, and (4) the number of negative field changes due to cloud flashes.

Table IV.—Field Changes due to Cloud Flashes in Near Storms.

Storm.	Distance.	Field changes.	
		Positive.	Negative.
	km.		
C ₂	4	16	3
J ₁	(OH)	3	1
R ₂	6	1	1
S	3	5	2
T ₂	4	8	0
W	4	2	0
AA ₂	4	3	6
AC	5	3	1
AD ₁	6	6	0
AD ₂	(OH)	10	1
AD ₃	2	2	1
AE ₁	3	3	0
AF ₂	4	2	0
AG	5	11	0
AH	4	3	0
AJ ₂	4	5	1
AM ₂	4	3	1
		86	18

The figures in this case are 86 positive and 18 negative field changes, giving a ratio of 4·8/1 in favour of the positive charge in the clouds being above the negative charge. This is evidence in favour of positive polarity but it is not so complete as to leave no doubt.

Schonland found at Somerset East that flashes occurring in storms nearer than 7 km. to the station, combining both cloud flashes and ground flashes, produced positive and negative field changes in the ratio 21/1. In Table IV only the cloud flashes are shown. If all the flashes to ground at a distance nearer than 6 km. to the station are added, the ratio, positive changes : negative changes becomes 8/1, which is just about half the ratio for Somerset East. The number of negative field changes produced by near ground flashes was only 3 out of 69. Thus the Johannesburg storms seem to give an unusual number of negative field changes for *cloud flashes* but not for ground flashes. In connection with this, one important factor must be taken into consideration.

In his second paper Schonland drew attention to the possibility of inclined flashes in the clouds producing complications and causing negative field changes nearer to the station than the simple theory of the thundercloud would indicate.* This led him to place the maximum distance at which cloud flashes would always give positive field changes at 7 km. Observations in Johannesburg have not only confirmed this suggestion but have indicated that the effect is of quite a large magnitude. At various times it has been noted by several "flash observers," including the writer, that flashes in the clouds are frequently inclined, sometimes almost reaching a horizontal position and often being of great length. Storm A was 30 km. away and quite a number of these inclined flashes subtended an angle at the station greater than 20° so that their length was of the order of 10 km.

During storm AD, sections (2) and (3), one observer reported that when standing in the quadrangle of the building he could see flashes crossing the sky vertically above him. These flashes did not reach the ground and must have been very inclined to the vertical to appear horizontal to an observer below them. At that time the interval between the lightning and the following thunder was 11 seconds, so that the height of the flashes was of the order of 4 km. Storm AA₂ shows 6 negative field changes at an estimated distance of 4 km. and in the records it appears that two consecutive flashes produced thunder after 24 seconds and 4·8 seconds respectively on one occasion, and after 11 seconds and 25 seconds respectively 20 minutes later. This gives

* 'Proc. Roy. Soc.,' A, vol. 118, p. 224 (1928).



FIG. 1.

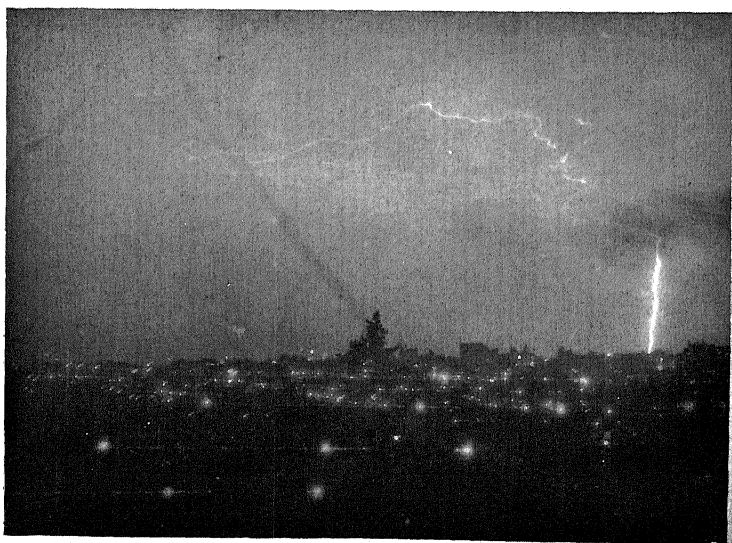
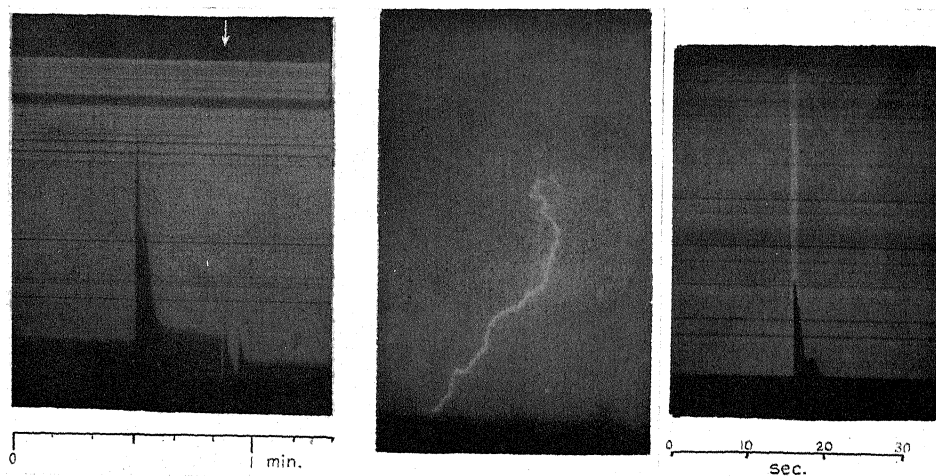


FIG. 2.



FIGS. 3 AND 4.

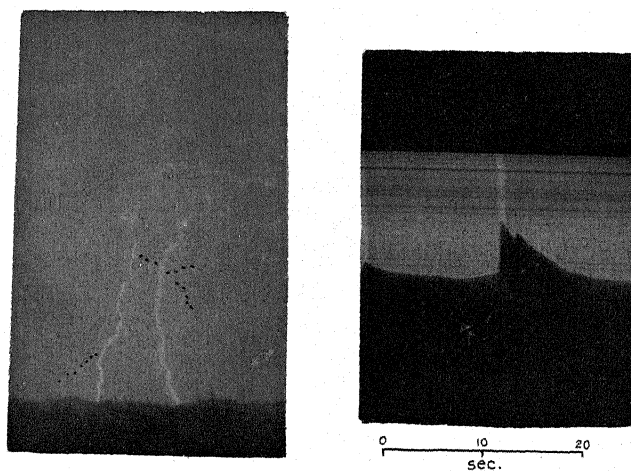


FIG. 5.

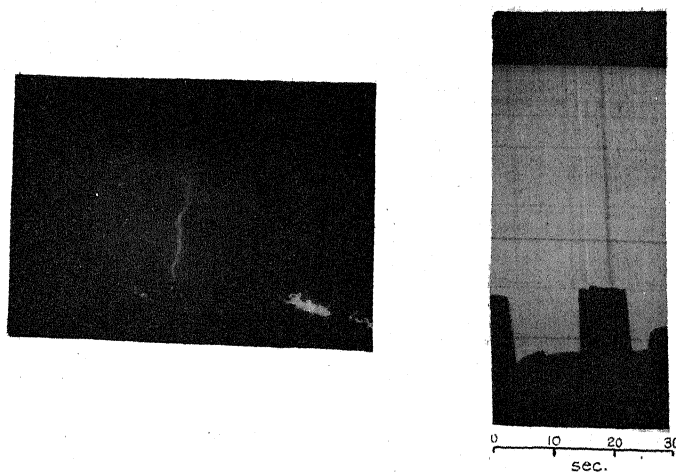


FIG. 6.

distances of 8 and 1.3 km. in the one case and 3.6 and 8.3 km. in the other, indicating that the active portion of the cloud was something like 6 km. across. Thus it is quite possible for the two poles of the cloud to be separated from each other by a horizontal distance of 6 km.

Fig. 1, Plate 5, shows a photograph of a flash in the clouds 5 km. away taken during storm AL on December 17, 1930. Its estimated length is 3 km. Unfortunately no record of the field change produced by this flash is available as the bubble of the electrometer had moved off the scale.

These instances make it clear that the observations upon flashes and field changes recorded above in Table IV are subject to quite important complications. Consider a flash in the clouds taking place from a positive pole 6 km. high to a negative pole 2 km. high, the positive pole being right overhead and the negative pole being distant 4 km. horizontally from the station. Such figures are substantiated by the figures recorded above. Then Wilson's expression for the field change as modified by Schonland (*loc. cit.*, vol. 114) can be employed, giving "L" two values, one for the positive pole and one for the negative pole.

$$\Delta F (\text{cloud}) = -2Q \left[\frac{H_2}{(H_2^2 + L_2^2)^{3/2}} - \frac{H_1}{(H_1^2 + L_1^2)^{3/2}} \right]$$

where H_2 and L_2 , H_1 and L_1 , are the height and horizontal distance respectively of the positive pole and the negative pole. Taking the figures suggested we have :—

$$\begin{aligned} \Delta F (\text{cloud}) &= -2Q \left[\frac{6}{(6^2 + 0)^{3/2}} - \frac{2}{(2^2 + 4^2)^{3/2}} \right] \\ &= -2Q(0.0278 - 0.0224) \\ &= -0.0108Q. \end{aligned}$$

The distance of that flash may easily be estimated as 4 km. If the flash is seen it is noted as (OH) in the log book and the field change produced by it is negative. It will be seen that only a special type of flash overhead will give rise to such a negative field change, but considering the large number of very inclined flashes seen from time to time this factor must be taken into consideration as being a possible cause of some of the negative field changes recorded during very near storms. Thus direct observation of lightning flashes and the occurrence of negative field changes occasionally with near flashes in the clouds leads to the general conclusion that in thunderstorms near Johannesburg, local conditions favour a lateral displacement of the two poles of the cloud relatively to each other.

7. Multiple Lightning Flashes.

During the progress of a storm it is sometimes noticed that a lightning flash consists of two discharges, one in the cloud and one from the cloud to the earth, with only a very short interval of time between them. Fig. 2, Plate 5, shows such a double lightning flash. An examination of the records of field changes shows an occasional case where a negative change is immediately followed by a positive change without any recovery between the two changes. Such a double change is shown in fig. 3, Plate 6. In Tables V, VI, and VII an analysis has been made of all the double field changes recorded during all the storms which have been observed. Table V shows all the double changes of the negative-positive type. Table VI contains all the positive-negative field changes, and Table VII includes positive-positive and negative-negative field changes. In Tables V and VI the third and fourth columns show whether the double field change was accompanied by a cloud flash or a flash to ground, while the fifth column contains those changes which were not accompanied by observation of the flash. The double field changes in Table VII were all unaccompanied by observation of the flash causing them.

Table V.—Negative-positive Field Changes.

Storm.	Distance.	Observed to be in cloud.	Observed to be to ground.	No observation of flash.
	km.			
Q	?	0	0	1
R	8	1	0	0
T	5	0	0	1
W	4	0	0	2
Y	7	0	1	0
AA	4	1	0	0
AB	15	0	1	0
AD ₃	2	2	0	0
AI	17	0	1	1
AJ	8	0	0	1
AL	3	1	0	1
AM	8 to 4	2	5	2
		7	8	9

Total number of field changes, 24.

An examination of the tables shows that double field changes have been associated equally with flashes in the cloud and with flashes to the ground, and that the greater number of the double changes are of the negative-positive

Table VI.—Positive-negative Field Changes.

Storm.	Distance.	Observed to be in cloud.	Observed to be to ground.	No observation of flash.
	km.			
R	8	0	0	2
S	3	0	1	1
AC	5	2	0	0
AI	17	0	1	1
AM	8 to 4	0	0	1
		2	2	5

Total number of field changes, 9.

Table VII.

Storm.	Distance.	+ + field changes.	— — field changes.
	km.		
G	4	0	2
J	8	1	2
U ₂	4	2	0
		3	4

Total number of field changes, 7.

type, the positive-negative changes being less frequent and the positive-positive and negative-negative changes quite rare.

On the supposition that the cloud consists of two areas of charge, the one above the other with the negative charge below, a negative-positive field change will indicate a flash in the cloud followed by a flash to the ground taking place at a distance from the station, or having the cloud flash so inclined that it produces a negative field change even when the cloud is nearer than 6 km. A positive-negative field change will indicate a similar set of circumstances but with the flash in the cloud taking place after the flash to the ground. The 24 flashes in Table V compared with 9 flashes in Table VI show that the flash in the cloud takes place more easily than the flash to the ground, and agrees with observations both by Simpson and Schonland that flashes take place more frequently in the cloud than to the ground.

Double field changes of the positive-positive type can be explained as being due to double flashes in near storms where cloud flashes as well as flashes to ground produce positive field changes. The double change in storm "J"

8 km. away may possibly be due to this cause, in which case the distance was slightly overestimated, or it may be due to two separately charged lower sections of the cloud discharging one after the other. The negative-negative field changes can only be due to two discharges in the clouds, and this might also be caused by the lower charge on the cloud being in two separate sections, first the one and then the other discharging to the upper pole.

Though the last two suggestions are made only in a tentative way the existence of a double flash of lightning consisting of a flash in the cloud followed by a flash to the ground is well established. In his second paper Wilson* states that the capillary electrometer would record field changes separated by an interval of $1/10$ th second. Experiments carried out with the electrometer used in this research indicate that in the present apparatus this value must be taken as an outside limit. With a high speed of travel of the photographic plate, 1 mm. per second and a slit of width $1/50$ th mm. a theoretical resolving power of $1/50$ th second is obtained but the limitations of the photographic plate do not allow this in practice. To obtain such a high rate of travel of the plate the emulsion must be a high speed one and this means large grains of silver so that the image is diffused. Besides this, in the apparatus used for this research, it is not possible to place the slit very close to the emulsion of the plate because space has to be allowed for the slide of the dark slide which contains the plate and therefore there is a spreading of the image which makes the minimum interval of time which can be recorded just about $1/10$ th second. With a slower travel of plate this interval is, of course, increased. With a travel of 1 mm. per second the plate makes a complete record in 1 minute and often it is not advisable or possible to make records at such a speed. Then the resolving power of the apparatus is diminished. Such being the case it is obvious that if the two flashes of lightning take place with an interval of time between them less than $1/10$ th second the field change recorded by the electrometer will be a single one, the resultant of the two separate field changes. Thus the record may show :—

- (1) A positive change, if the earth flash causes a bigger field change than the cloud flash.
- (2) No change, if the two flashes each produce the same magnitude of field change.
- (3) A negative change, if the effect of the cloud flash is greater than that of the flash to the ground.

* *Loc. cit.*, 'Phil. Trans.'

The effect which this will have in producing anomalous field changes during a storm must be considered. In the first place it must be remembered that, as has been mentioned in section 3, a flash to the ground is much more easily observed than a flash in the clouds. As a result the "flash observer" is more than likely to note a double flash as a "ground flash." Effect (1) is therefore noted as "ground flash—positive" and appears to be a normal case although it is not. Effect (3) is noted as "ground flash—negative" and appears to be evidence of negative polarity although it is really a special case of positive polarity. Effect (2) is noted as "ground flash—no change" and immediately shows that the flash is not a simple one. A photographic correlation of flashes will show all three effects clearly, for the photograph will reveal the double flashes as visual observation cannot. For example the flash shown in fig. 2 would certainly have been reported as a flash to the ground without the aid of the photograph. Table VIII gives the results of a few photographs of flashes which show the possibilities of the method, although sufficient have not been obtained to provide a weighty body of evidence.

Table VIII.—Double Flashes. (In Cloud and to Ground.) Photographic Correlations.

Field changes.		
Positive.	Negative.	Zero.
2	8	1

Information for the study of effect (2) has been obtained by visual observation of the flashes and is given in Table IX, which gives the number of flashes producing no field change which have been observed in the storms under discussion. The first column gives the index letter of the storm, the second column gives its distance from the station, the third column shows the number of flashes reported "cloud flash" and the fourth column shows the number of flashes reported "ground flash."

The table shows that 46 flashes of lightning to the ground produced no field change and it is very likely that these were really double flashes of which only the discharge to the ground was seen by the "flash observer."

The occurrence of flashes which produce no field change having been demonstrated, there is no reason to doubt that similar flashes took place as has already

Table IX.—Flashes Producing No Field Change.

Storm.	Distance.	Observed to be in cloud.	Observed to be to ground.
	km.		
R ₁	8	0	2
S	3	2	2
T	8	0	1
U	6	0	5
W	4	2	0
Y	11	0	2
AA	7	3	1
AB	13	0	1
AD	6	3	2
AF ₁	11	0	4
AG	5	2	1
AJ ₂	4	3	2
AK	12	2	5
AL	6	9	10
AM ₁	8	10	8
		34	46

been suggested, where the charges on the two poles were of such relative magnitudes as to make the nett resultant field change negative. This probably accounts for some of the 16 negative field changes due to flashes to the ground shown in Table III.

The point may quite conceivably be raised that perhaps the cloud is really of the nature suggested by Simpson (*loc. cit.*), with a positive charge located in the lower front portion of the cloud and a negative charge spread over the upper portion of it and extending to the rear, and that the positive field changes due to flashes to the ground are really due to double flashes of which only the flashes to the ground are seen. The answer to this is along three lines :—

(1) Such a flash as is under discussion is quite definitely not a normal flash. The most usual discharge is a single flash either between the two poles or from the lower pole to the earth and only special conditions will produce a second flash. Still more special conditions would be required to produce the two flashes so close together in time that they would be recorded as a single field change.

(2) The sign of the field change produced by a double flash depends upon the distance of the storm from the station as well as upon the relative magnitudes of the two charges, and in a table such as Table III where observations upon storms at all distances are recorded, it should be found that at certain distances positive field changes should predominate and at other distances negative field changes should predominate.

For both these reasons a cloud of the type suggested by Simpson could not be expected to give a large number of positive field changes. Table III shows that 267 positive and 16 negative field changes were observed to be caused by flashes to the ground. It is more logical to make use of the existence of a small number of special double flashes to explain the 16 negatives than to explain the 267 positives.

(3) Table VIII shows that out of 11 flashes photographed and shown to be double, 8 were accompanied by a negative field change. Thus the indications are that in the case of a double flash the field change, if it is a single change, is more likely to be negative than positive. This again shows that the abnormal field change is not the positive change but the negative one and supports the view that the lower charge on a thundercloud is a negative one.

The 34 cloud flashes producing no field change are also worth consideration. It is scarcely possible that they were all double flashes of which the observer saw only the cloud flash and missed the ground flash. It is more probable that they were cloud flashes taking place in the reversal zone where the field change due to a cloud flash is zero.

In the 4 storms at a distance greater than 8 km. from the station only 2 flashes were seen which produced no field change. In the 4 storms 7 and 8 km. away from the station, 3 storms supplied 3 flashes producing no change, and the remaining storm is "AM" whose distance could not be definitely determined (the frequency of the flashes made the timing of the thunder interval very difficult) and which produced 10 flashes in the cloud without any field change. The remaining 19 flashes all occurred in storms closer to the station than 7 km. Ruling out the doubtful storm we have:—

4 storms more than 10 km. away—2 flashes with no field change.

3 storms 7 and 8 km. away—3 flashes with no field change.

7 storms less than 7 km. away—19 flashes with no field change.

Thus the reversal zone for the field change produced by cloud flashes seems to start at a distance from the station of about 7 km. This confirms the remark that the reversal distance at Johannesburg is less than at Somerset East (§ 6 (a)) and justifies the inclusion of storms 9 km. away in the table of distant storms.

8. Branched Lightning Flashes.

Many of the photographs of lightning flashes taken in Johannesburg for the purpose of making correlations, show branching. On the suggestion of

Dr. Schonland these photographs have been examined, together with the records of field changes made at the same time, and the results obtained are given in Table X. This table contains all the photographs which showed flashes to ground only, double flashes not being included. The first three columns give the number of flashes with downward branching according as they produced positive or negative field changes or where no record of the field change was obtained. The next three columns show the unbranched flashes divided according to the same rule.

Table X.

Flashes branched down.			Flashes unbranched.		
ΔF positive.	ΔF negative.	No record of fields.	ΔF positive.	ΔF negative.	No record of fields.
10	0	8	6	0	1

None of the flashes showed upward branching, and in none of the photographs did a downward branching flash produce a negative field change.

The ratio of *branched* flashes to *unbranched* flashes is 18/7 or 3/1. This is liable to be an underestimation of the value because with increasing distance from the station the intensity of the flashes falls off and the branches, which are not as bright as the flashes, do not show on the photograph. The ratio positive/negative for flashes to ground is 16/1. It is therefore inevitable that since the branched flashes outnumber the unbranched flashes and the positive field changes outnumber the negative field changes, then negative electricity must pass to the ground by way of flashes which have downward branches. The eight branched flashes where the field changes were not observed can be treated statistically. Since all the observed branched flashes produced positive field changes it cannot be assumed that all the unobserved flashes produced negative field changes. It is more likely that at least half of them produced positive field changes. Therefore it can be taken that out of 18 branched flashes 14 produced positive field changes and only 4 produced negative field changes. This also indicates that it is possible for lightning flashes to fork from a negative charge to a positive charge. It is therefore concluded that the forking of a flash does not give an indication of its direction or the polarity of the part of the cloud from which it started. This conclusion has recently been subjected to laboratory tests by Schonland and Allibone.*

* 'Nature,' vol. 128, p. 794 (1931).

9. Some Photographic Data.

In this section some examples are given of the photographic data obtained during the investigation to illustrate the arguments in the previous sections of the discussion.

Illustrations are given of the following types of correlation :—

- (1) A distant flash to ground with a positive field change.
- (2) A near flash to ground with a positive field change.
- (3) A distant flash in the cloud with a negative field change.
- (4) A near flash in the cloud with a positive field change.
- (5) A double flash of lightning with a positive field change.
- (6) A double flash of lightning with a negative field change.

Wherever possible the record of the field change has been shown with the photograph of the flash. In a few cases the record of the field change has been too thin to make a clear reproduction and so the sign of the change is noted in the text. Where reproductions of the field change records are shown, a movement of the mercury meniscus, as shown by the black shadow, in an upward direction indicates a positive change of field. In each case a time scale is marked under the record in units of 10 seconds. The instant when a flash of lightning was seen is marked on the field records by a white line caused by brightening the light. This causes the image of the slit of the recording apparatus to spread a little owing to a certain amount of halation produced by the bright light, and in some cases, where the key was pressed by the "flash observer" very soon after the flash was seen, it caused the bright line on the record to extend to the left of the sudden mercury movement. This was very likely to happen when the rate of movement of the photographic plate was slow, and in some of the illustrations the effect is clearly seen. It does not mean that the indicator mark was made before the field change was produced.

Photographic Correlation 1. Fig. 4, Plate 6.—This photograph, which was taken during storm AL, shows a flash to earth which produced a positive field change. Its distance could not be estimated as it was very difficult to time the thunder interval owing to the frequency of the flashes, but it was not very near and the photograph gives the impression that it might have been more than 7 km. distant. The apparent turning back of the flash upon itself may be an illusion. It is possible that the flash had a horizontal portion at the top as can be seen in other photographs, and what is seen is a foreshortened view

of this horizontal portion. It may even be that the photograph really shows a cloud flash and a ground flash, in which case it is an example of a double flash producing a positive field change. The flash is not a very near one, and comes to ground without branching. The field produced by the storm was negative, and the field change produced by the flash was positive.

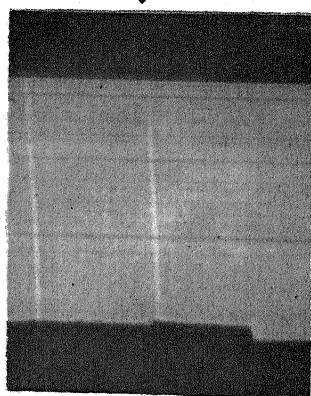
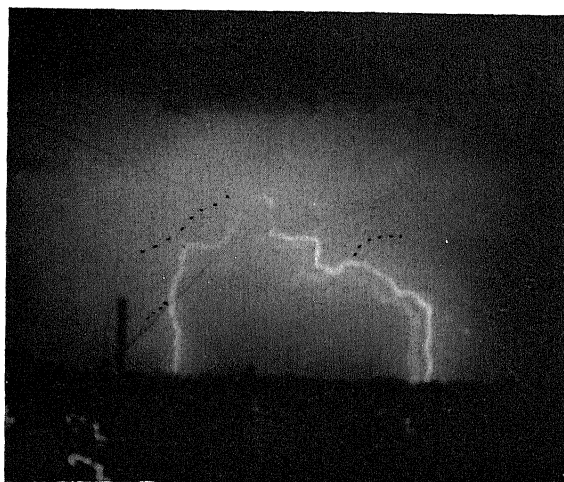
Photographic Correlation 2. Fig. 5.—This photograph was also taken during storm AL and seems to be at about the same distance as the other. The forks on the left-hand flash, which can be seen clearly on the negative, are very thin and so have been indicated by dots above them. The field change record shows a second positive change taking place 1·5 seconds after the one marked. There is a possibility that this second change was produced by one of the two flashes shown in the photograph, for when the camera was on a stand the “flash observer” generally pressed the key for marking the field record before closing the camera shutter. The flash is branched downwards to the ground and was not very near to the station. The field change produced by the flash was positive.

Photographic Correlation 3. Fig. 6.—This flash was photographed during a storm which occurred during the latter part of 1931 and which is not included in the storms discussed above. It provided some photographic correlations which are included here as illustrations. Its catalogue number is CG. This flash was about 20 km. away from the station. As a result it is impossible to tell if it was branched or not. It struck the ground, for the shadow of a tree and the horizon can be seen just at its foot. The flash was a distant flash to the ground and the field was a small positive one—about the normal fine weather field. The field change produced by the flash was positive.

Photographic Correlation 4. Fig. 7, Plate 7.—This photograph was also taken during storm CG. The branches on the flashes which are very faint have been indicated by dots near to them. The flash was a downward branching flash to ground, and the field change produced by the flash was positive.

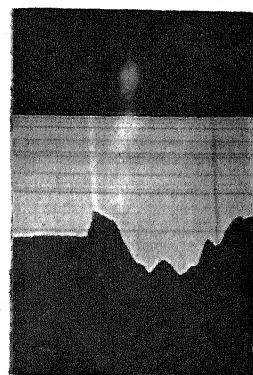
Photographic Correlation 5. Fig. 8.—This photograph, also taken during storm CG, shows a flash to the ground taking place at a distance about 20 km. away. No details of forking can be seen. The field change produced by the flash was positive.

Photographic Correlation 6. Fig. 9, Plate 8.—This photograph was taken during storm AM and shows a flash which caused a voltage surge and “flash over” on a high-tension supply wire in the city so that it must have been very near. The indicator mark on the record is 2·2 seconds after the field change produced by the flash because the blaze of light in the city caused the observer



0 10 20
Sec.

FIG. 7.



0 10 20
Sec.

FIG. 8.

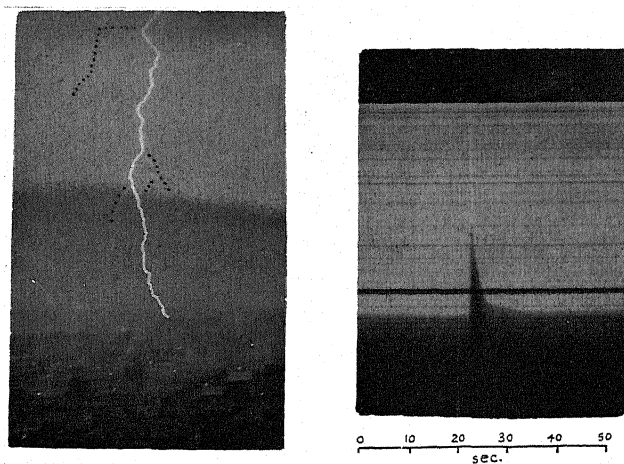


FIG. 9.

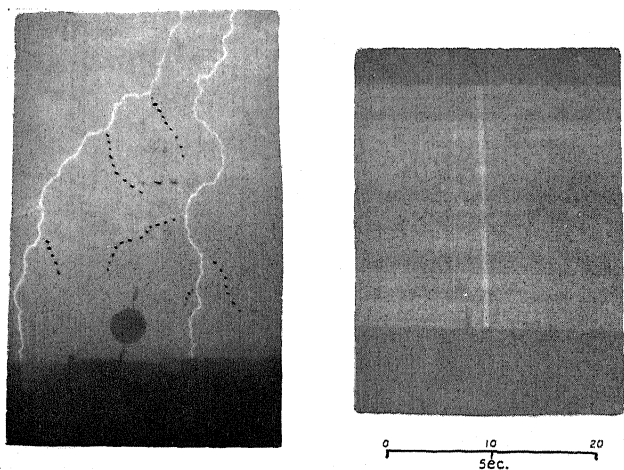


FIG. 10.

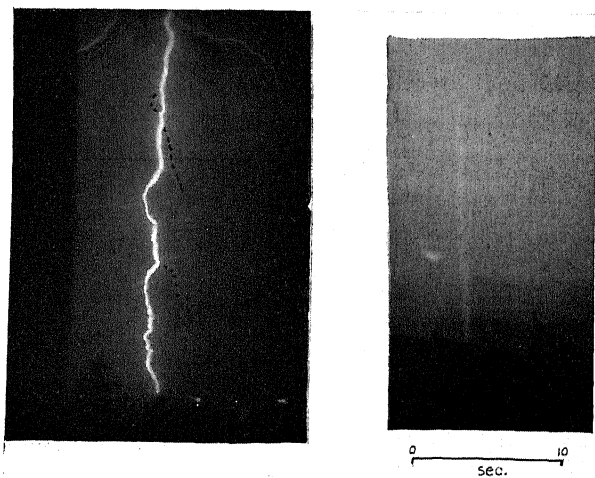


FIG. 11.

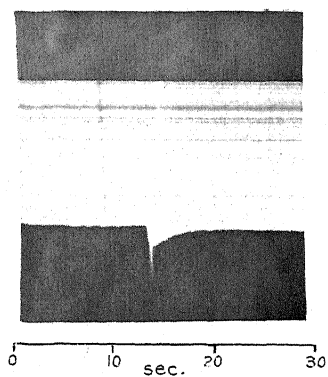


FIG. 12.



FIG. 13.

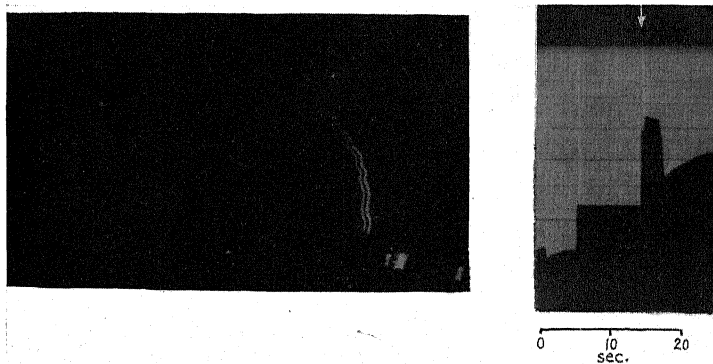


FIG. 14.

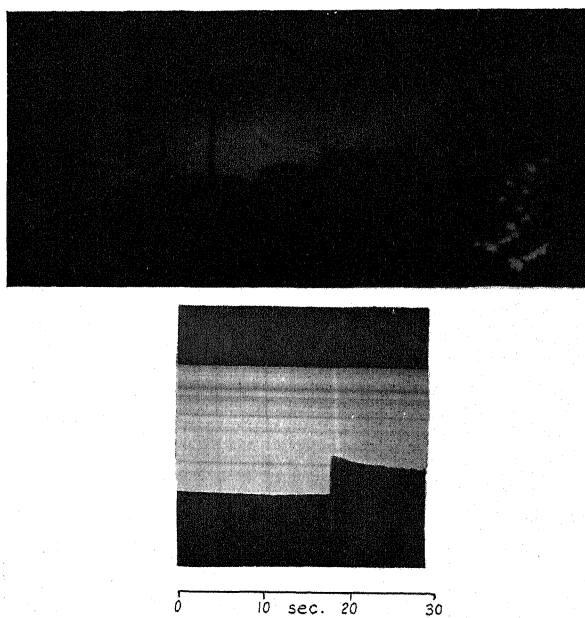


FIG. 15.

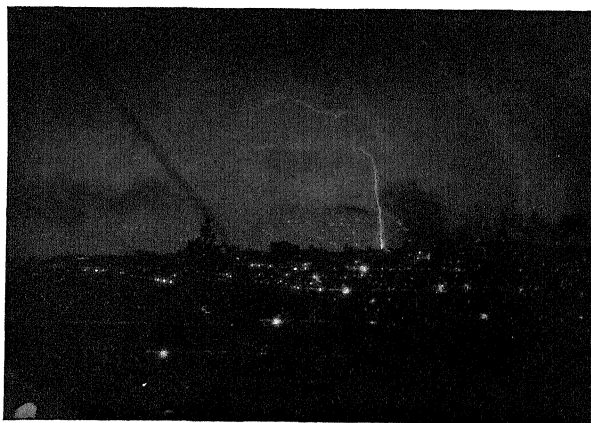


FIG. 16.

to forget to press the marking key at once. The flash branched downwards to the ground. Its distance was about 2 km. The field produced by the storm was negative, while the field change produced by the flash was positive.

Photographic Correlation 7. Fig. 10.—This flash was photographed during the same storm (AM) 3 minutes after flash number 6. Its distance was estimated as 3 km. and it produced a positive field change. Downward branching can be seen in the photograph and the branches have again been indicated by dots. The ball in the foreground is the copper ball which was being used to measure the field change.

Photographic Correlation 8. Fig. 11.—This flash (storm AD) was observed to produce thunder after 13 seconds corresponding to a distance of 4.3 km. from the station. It will be seen in the photograph that it struck the ground behind a mine dump which is known to be over 4 km. distant, so that the estimated distance may be too small by several kilometres. This illustrates another way in which photography is of use in obtaining reliable information about a storm. The flash is a downward branching flash to ground. The distance of the flash was about 6 km. The field produced by the storm was negative and the field change produced by the flash was positive.

Photographic Correlation 9. Fig. 12, Plate 9.—This flash, which was photographed during storm CG, was a cloud flash taking place at a distance. It will be seen that the flash is almost horizontal, emerging from the cloud in the centre of the picture, passing into the cloud again farther to the right and then emerging again right at the edge of the field of view. Its horizontal dimensions are about 5 km. It illustrates the distorted nature of some of the thunderclouds which are observed in Johannesburg. The flash was a distant cloud flash, about 20 km. The field change produced by the flash was negative. The white line between the factory chimney and the cloud flash is not a flash to ground but a mark on the plate.

Photographic Correlation 10. Fig. 13.—This photograph shows a flash in the clouds at a distance of 2 km. from the station. It will be noticed again how very nearly horizontal it is. At the bottom edge of the picture a tree and a large light outside a cinema indicate the position of the ground. In the top left-hand corner is the silhouette of one of the insulators of the aerial used for recording distant storms. The field change was positive but the record is not clear enough to be reproduced.

Photographic Correlation 11. Fig. 14, Plate 10.—This photograph shows a double flash of lightning. The cloud flash, as is often the case, is very weak owing to the screening effect of the clouds, but is quite visible on the negative. The

second field change on the record is the one which was produced by the double flash. This is a case of a double flash producing a positive field change. The flash took place in storm CG, over 20 km. away. It is probable that the very inclined nature of the cloud flash caused the field change from it to be small in comparison with the positive change due to the flash to the ground.

Photographic Correlation 12. Fig. 15.—This flash, also taken during storm CG, shows a very long cloud flash, almost horizontal, which at one point sends a flash to the ground thus constituting a double flash. It touches ground to the left of the factory chimney and also sends branches to the ground to the right of the chimney. These branches cannot be seen in the reproduction. This is another case of a double flash producing a positive field change.

Figs. 2 and 16 show double flashes which produced negative field changes. The photographs were taken in 1929 before records of the field changes were made photographically, so the field changes cannot be shown. The distance of the two flashes was about 8 km.

In conclusion the author wishes to thank Dr. B. F. J. Schonland of the Physics Department, University of Cape Town, at whose suggestion the work was undertaken and who has provided much valuable advice and assistance ; Professor H. H. Paine of the Physics Department, University of the Witwatersrand, for many suggestions and the use of apparatus, and the Research Grant Board of the Union of South Africa for a grant in aid.

10. *Summary.*

An account has been given of a method of recording photographically the sudden changes of electric field caused by a flash of lightning and at the same time of observing the flash visually, photographing it if possible and making a mark on the field change record immediately after it took place.

An analysis has been made of 560 correlations of "nature of flash" with "sudden field change" made in this way, on the lines indicated by a study of the effects to be expected from lightning flashes taking place from clouds where the positive charge of electricity was above the negative charge, and following on the work done by Wilson and Schonland on this subject. It has been shown that the field changes obtained when flashes of lightning took place from the cloud to the ground and when they took place in the cloud when it was at a distance from the observing station were such as to support the view that the upper charge on a thundercloud is positive and the lower charge is negative.

The occurrence of "double" lightning flashes has been discussed, together with the electrical effects which such a compound flash would produce.

The question of the exact significance of branching in a lightning flash has been investigated, and it has been shown that negative electricity can travel from the cloud to the ground by way of a flash which branches towards the ground.

The conclusions arrived at have been supported by photographs of lightning flashes illustrating the points considered.

Inelastic Electron Scattering in Gases.—I.

By C. B. O. MOHR, B.A., M.Sc., Trinity College, Cambridge, 1851 Exhibitioner, University of Melbourne; and F. H. NICOLL, M.Sc., Trinity College, Cambridge, 1851 Exhibitioner, University of Saskatchewan.

(Communicated by Lord Rutherford, O.M., F.R.S.—Received June 25, 1932).

Introduction.

The study of the scattering of electrons in gases has been considerably extended during the last three or four years by numerous investigators, and the results in a few cases have recently been interpreted with some success by the wave mechanics.

The earlier experiments of Dymond and Watson,* Harnwell,† and McMillen‡ on the angular distribution of the scattered electrons were confined to electron velocities above 50 volts and to angles of scattering smaller than 60° ; the results showed only a steady decrease in the scattered intensity with increasing angle of scattering, the steepness of the curves increasing with increased electron velocity.

More recently the elastic scattering in various gases has been systematically investigated in this laboratory by Bullard and Massey,§ and by Arnot,|| and the measurements carried out over a large range of electron velocities and an

* 'Proc. Roy. Soc.,' A, vol. 122, p. 571 (1929).

† 'Phys. Rev.,' vol. 34, p. 661 (1929).

‡ 'Phys. Rev.,' vol. 36, p. 1034 (1930).

§ 'Proc. Roy. Soc.,' A, vol. 130, p. 579 (1931); vol. 133, p. 637 (1931).

|| 'Proc. Roy. Soc.,' A, vol. 130, p. 654 (1931); vol. 133, p. 615 (1931).

angular range of from 15° to 125° . In all cases diffraction effects were found to occur, maxima and minima appearing at large angles. In certain cases, theoretical curves have been obtained which agree fairly closely with the experimental.* These results have since been verified and extended to larger angles of scattering and lower velocities by Ramsauer and Kollath† in various gases, by Pearson and Arnquist‡ in mercury vapour, and by Hughes and McMillen§ in argon.

More recent investigations of the inelastic scattering have been carried out by Rose|| in mercury vapour, and by Hughes and McMillen¶ in argon, the observations again being confined to angles of scattering of less than about 50° . The results again showed only a steady decrease in the scattered intensity with increased angle of scattering, this decrease being more rapid than in the case of the elastic scattering for electrons of the same incident velocity.

The present paper describes experiments on the angular distribution of the inelastic scattering for the case of electrons which have lost energy in raising the atom to the most probable excited state. The investigations have been carried out in helium, argon and mercury vapour over an angular range of from 20° to 160° , using electrons with incident energies between 23 and 196 volts. In the case of argon and mercury vapour, the curves obtained show maxima and minima at large angles, and are markedly similar in most cases to the corresponding curves for the elastic scattering. Incidentally, good agreement is obtained with the results of previous observers for the elastic scattering, the measurements being extended to an angle of 160° .

Description of the Apparatus.

The apparatus used is illustrated diagrammatically in fig. 1A. It consisted essentially of a cylindrical chamber C, in which the electron collisions occurred, and a chamber A in which the scattered electrons were analysed.

Electrons were fired towards the centre O of the collision chamber from the electron gun G, which was rotated about the axis O by means of a ground glass joint; the gun consisted of a straight tungsten filament surrounded by an outer case containing a pair of slits $1.1 \text{ mm.} \times 6 \text{ mm.}$, and 6 mm. apart;

* *Vide* Bullard and Massey, 'Proc. Roy. Soc.,' A, vol. 133, p. 637 (1931).

† 'Ann. Physik,' vol. 12, pp. 529, 837 (1932).

‡ 'Phys. Rev.,' vol. 37, p. 970 (1931).

§ 'Phys. Rev.,' vol. 39, p. 585 (1932).

|| 'Canadian J. Res.,' vol. 3, p. 174 (1930).

¶ 'Phys. Rev.,' vol. 39, p. 585 (1932).

the magnetic field of the filament was reduced to a minimum by placing the return lead as close as possible to the filament.

Electrons scattered in a small region about O passed through two slits S_1 (0.55×3.4 mm.) and S_2 (0.24×2.8 mm.), 10 mm. apart, into the analyser A, in which the electrons of different energies were sorted out by a uniform radial electrostatic field.* The electrons then passed into the Faraday cylinder F, enclosed in a case carrying a slit 0.2 mm. \times 2.8 mm.

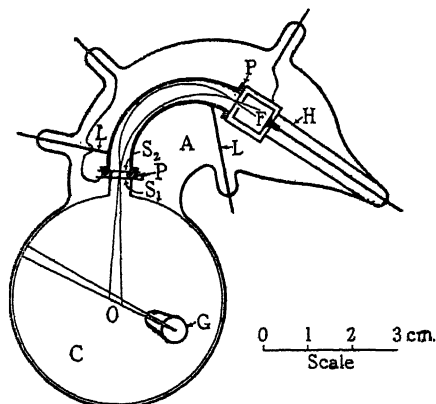


FIG. 1A.—Diagram of apparatus.

The collision chamber was lined with thin metal in order to produce a field-free space, but near the analyser A; the metal lining was shaped so as to prevent electrons passing into the analysing chamber or analyser except through the slits S_1 and S_2 .

The deflecting plates of the analyser A were made of copper, accurately bent to the required shape; they were fixed to the slits S_1 and S_2 and to the Faraday cylinder case by means of small copper screws, insulated by small pieces of steatite P, previously cut to the required shape and baked hard. The Faraday cylinder was insulated by a piece of quartz tubing, and the lead to it shielded by a light metal cylinder H. This analysing system was supported in position, with S_1 , S_2 , and O in a straight line, by means of two rigid copper wires L, L_1 , silver-soldered to tungsten leads, the latter being sealed into the glass side-tube containing the analyser.

The other metal parts were made of copper or Eureka, and the whole system was enclosed inside a Pyrex vessel. The whole apparatus, with the exception of the ground joint, could thus be baked at a temperature of about 450° C.,

* Hughes and Rojansky, 'Phys. Rev.', vol. 34, p. 284 (1929). Rudberg, 'Proc. Roy. Soc.,' A, vol. 129, p. 628 (1930). Van Atta, 'Phys. Rev.', vol. 38, p. 876 (1931).

the apparatus, in fact, being designed with this end in view. Without this precaution, stray scattering from gas and films of grease on slits, etc., was obtained which, even at small angles, was comparable with the scattering by the gas, and at large angles completely obscured it.

The apparatus was connected through a liquid air trap to a mercury diffusion pump backed by a Hyvac oil pump, and the pressure of the residual gases inside the apparatus could be maintained below 10^{-5} mm. Hg. When investigating the scattering in mercury vapour, liquid air was not placed on the trap, and mercury vapour was allowed to diffuse over from the mercury diffusion pump. Before investigating the scattering in helium and argon, the pressure of the mercury vapour in the apparatus was reduced to a negligible quantity by prolonged baking, liquid air being kept on the trap; the gas to be used was then passed from a reservoir through a fine leak into the apparatus, the pressure of the gas being maintained at a value less than 10^{-3} mm. Hg.

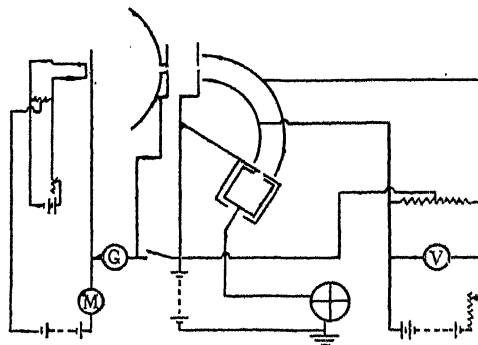


FIG. 1B.—Electrical connections.

The electrical connections are illustrated in fig. 1B. The slits at each end of the analyser A were kept at a potential midway between the potentials of the two deflecting plates by means of two equal high resistances as shown, a potentiometer arrangement being used to apply a variable potential between the deflecting plates. A retarding potential was placed between the Faraday cylinder and its case in order to reject all electrons of lower velocity than those it was desired to collect, and so remove any possible background due to slow electrons. A potential could be applied between the slits S_1 and S_2 in order to accelerate the scattered electrons at the lower voltages, if desired, before entering the analyser; actually it was not found necessary to do this. The filament emission was measured by a milliammeter M, the electron beam leaving the gun by a galvanometer G, and the voltage applied to the analyser plates by an

accurate voltmeter V; the current collected by the Faraday cylinder was measured by the rate of drift of a Dolezalek electrometer.

The apparatus, when first constructed, consisted of an analysing chamber, separated from the scattering chamber by a glass inner seal carrying a small slit $0.1 \text{ mm.} \times 3 \text{ mm.}$, and the analyser was connected by a piece of short wide pyrex tubing to a large liquid air trap. Thus, while the scattering chamber contained, say, mercury vapour at a pressure corresponding to room temperature, the pressure in the analyser could be kept much lower. Also the analyser plates were of radii 3.5 and 5 cm. respectively, and 4 cm. wide. After several curves had been obtained for the angular distribution of the inelastic scattering in mercury vapour, using this apparatus, it was decided, as a result of the experience gained, to modify it slightly in order to effect several improvements. For example, it was found that, if the length of path could be decreased slightly in the analyser, the effect of not evacuating the analyser was of little importance, and caused no appreciable difference in the results obtained. Without the necessity for evacuating the analyser, the design could be much simplified, and the problem of alignment of the various slits rendered much easier. The apparatus in its second form was as described at the beginning of this section, and the analyser plates were of radii 1.3 cm. and 1.9 cm. respectively, and of width 1.5 cm. The measurements of angular distributions in mercury vapour were repeated with this apparatus, and found to agree closely with those obtained with the earlier apparatus.

Experimental Procedure.

Before measuring the angular distribution of the scattered electrons it was first necessary to investigate their energy distribution by measuring the collected current over a range of deflecting voltages. Peaks appeared in the energy distribution corresponding to the elastically and inelastically scattered electrons, the latter having energies corresponding to various energy losses, while the value of the deflecting voltage required to focus the elastically scattered electrons was found to agree closely with that calculated from the dimensions of the analyser.

With the voltage on the deflecting plates fixed at the value corresponding to the top of a peak, the angle of scattering was varied by rotating the ground joint carrying the electron gun, and the scattered intensity obtained as a function of angle. The zero of angle was obtained from the symmetry of curves taken for positive and negative angles in a preliminary run. The

electron beam current, as measured by the galvanometer, did not vary by more than 10 per cent. during a run, and the scattered current was corrected for this variation by dividing each reading of the collected current by the galvanometer reading.

That the height of the peak was a measure of the number of electrons scattered was verified by comparison of energy distribution curves taken at a large number of angles, when it was seen that the width of the base of the peak remained constant, so that the area under the peak was proportional to the peak height. A correction for the change of scattering volume with angle of scattering was made, in the usual manner, by multiplying the scattered intensity by the sine of the angle of scattering.

In preliminary experiments the following checks were made on the working of the apparatus :—

(i) With no gas or vapour in the apparatus, no measurable scattered current was collected by the Faraday cylinder except at angles less than about 40° , when it was only a few per cent. of that obtained with gas in the apparatus ; at angles less than about 4° the main electron beam from the gun began to enter the analyser.

(ii) The collected current was proportional to the total current from the gun as measured by the galvanometer.

(iii) The results obtained were shown to be due only to single scattering in the gas, as follows : the variation of the scattered current with pressure was investigated for several different angles of scattering greater than 20° , and found to be linear over the range of pressures used in the experiments ; some of the results obtained for helium and argon are shown in fig. 2. At higher pressures than that shown in the figure, a departure from linearity began to set in, probably owing to collisions in the analyser as much as to the onset of multiple scattering in the scattering chamber. In the case of mercury vapour it was necessary to measure the pressure with an ionisation gauge, and to alter the pressure of the vapour by heating or cooling the apparatus ; although the results were not as accurate as might have been desired, owing to the difficulty of maintaining uniform and steady temperature conditions, there seemed to be no marked deviation from single scattering for pressures below 10^{-3} mm. Hg, the pressure of mercury vapour at room temperature.

(iv) That the scattered electrons collected by the Faraday cylinder really had the energy expected, from a knowledge of the potential applied to the focussing plates, was checked by varying the retarding potential between the Faraday cylinder and its case, and so obtaining retarding potential curves.

(v) The results obtained for the elastic scattering were found to agree closely with those obtained by Bullard and Massey, and Arnot (*loc. cit.*) with a different type of apparatus.

We now proceed to consider the results obtained for the scattering in the gases investigated.

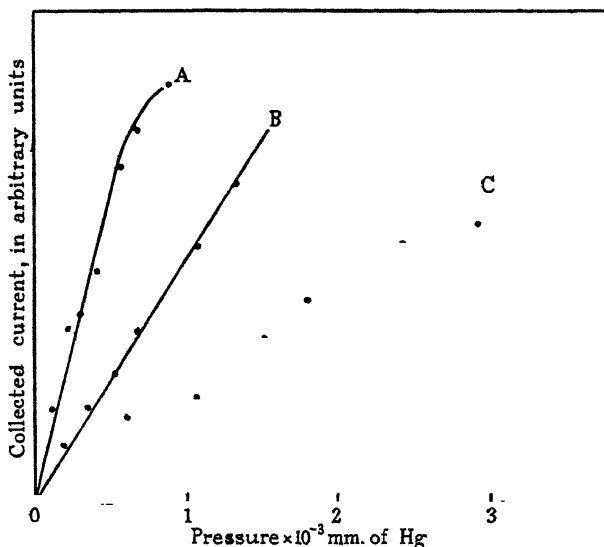


FIG. 2.—Illustrating ratio of scattered current to pressure.

A.—42 volts, inelastic, 40° Argon.

B.—84 volts, inelastic, 30° Helium.

C.—62.5 volts, elastic, 30° Helium.

Results.

(a) *Helium*.—Typical curves showing the energy distribution near the main inelastic losses in the three gases investigated are given in fig. 3; the curves are plotted with the energy loss as abscissa.

The curve for helium is for incident electrons of 42 volts energy and an angle of scattering of 30°. In view of the work of Whiddington and Roberts* and Van Atta,† it is clear that the peaks at about 21 volts and 23 volts correspond respectively to the excitation of the helium atom to the 2^1P and 3^1P levels (21.11 volts and 22.96 volts), while there are indications of the 23.6-volt loss corresponding to excitation to the 4^1P level.

* 'Proc. Leeds Phil. Soc.', vol. 2, p. 201 (1931).

† 'Phys. Rev.', vol. 38, p. 876 (1931).

In fig. 4 are plotted points showing the angular distribution of the elastic and inelastic scattering (21.11 volts loss) for electrons of 54, 83, 120 and 196 volts energy, and for angles up to 60° ; the angular distribution for the case of the 22.96-volt loss for 83-volt electrons is also given. In order to compare the results with theory, the figure shows calculated curves obtained by the use of the ordinary Born formula; the question of agreement with theory will be discussed later in the paper.

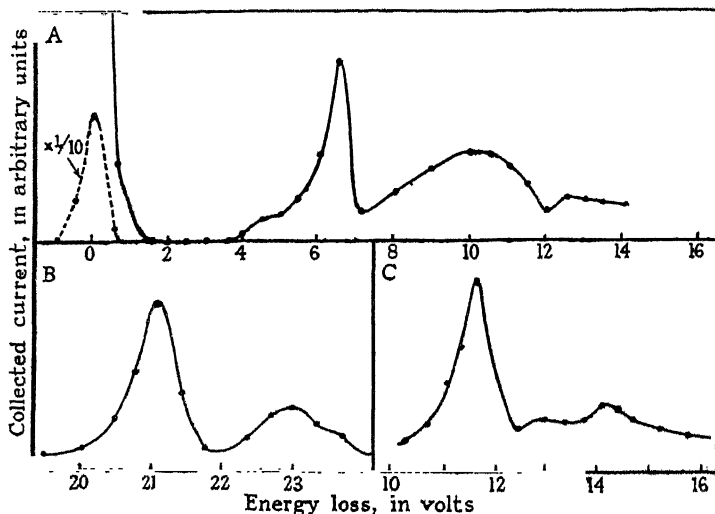


FIG. 3.—Energy losses in Mercury Vapour, Helium and Argon.

A.—Mercury Vapour, 90° , 42 volts.

B.—Helium, 30° , 42 volts.

C.—Argon, 20° , 42 volts.

In order to show the large angle scattering, the curves have been re-plotted on a different scale in fig. 5, and extended to an angle of 160° .

(b) *Argon*.—The main inelastic loss in argon corresponds to an energy loss of 11.6 volts, as found by Van Atta, and Hughes and McMillen (*loc. cit.*). A typical energy distribution curve in argon is shown in fig. 3 for incident electrons of 42 volts energy and a scattering angle of 20° .

The angular distribution of the elastic and inelastic scattering (11.6 volt loss) in argon for incident electrons of 42 volts and 61 volts energy is illustrated in fig. 6; the curves are arbitrarily fitted together at one point. It was, unfortunately, not possible to measure the large angle scattering at lower voltages owing to lack of intensity, while at higher voltages the losses near the ionisation potential became more prominent than the 11.6 volt loss, and were not completely separated from the latter by the analyser.

It is seen that the inelastic curves exhibit maxima and minima at about the same angle as the elastic curves, although the maximum is much less pronounced, particularly in the case of the 61-volt curves.

(c) *Mercury Vapour*.—A typical energy distribution in mercury vapour is shown in fig. 3 for incident electrons of 42 volts energy, and a scattering angle

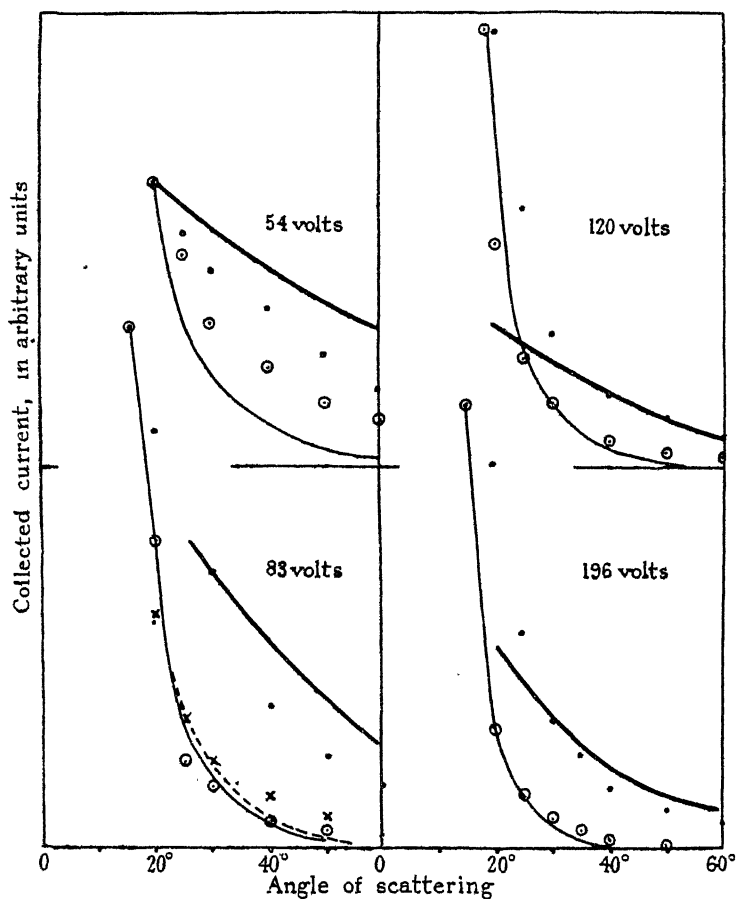


FIG. 4.—Small angle scattering in Helium.

- • Experimental points, elastic — Theoretical elastic.
- ⊙ ⊙ Experimental points, inelastic [$2^1 P$] — Theoretical $2^1 P$.
- × × Experimental points, inelastic [$3^1 P$] - - - Theoretical $3^1 P$.

of 90° . On the left is shown the elastic peak on one-tenth of the scale; the main peak at 6.7 volts is due to excitation to the 1^1P_1 level, while there are indications of a peak at 4.9 volts corresponding to excitation to the 1^3P levels.

Above 8 volts are grouped together, unresolved, the various energy losses near the ionisation potential obtained by Whitney,* Foard† and Rose.‡

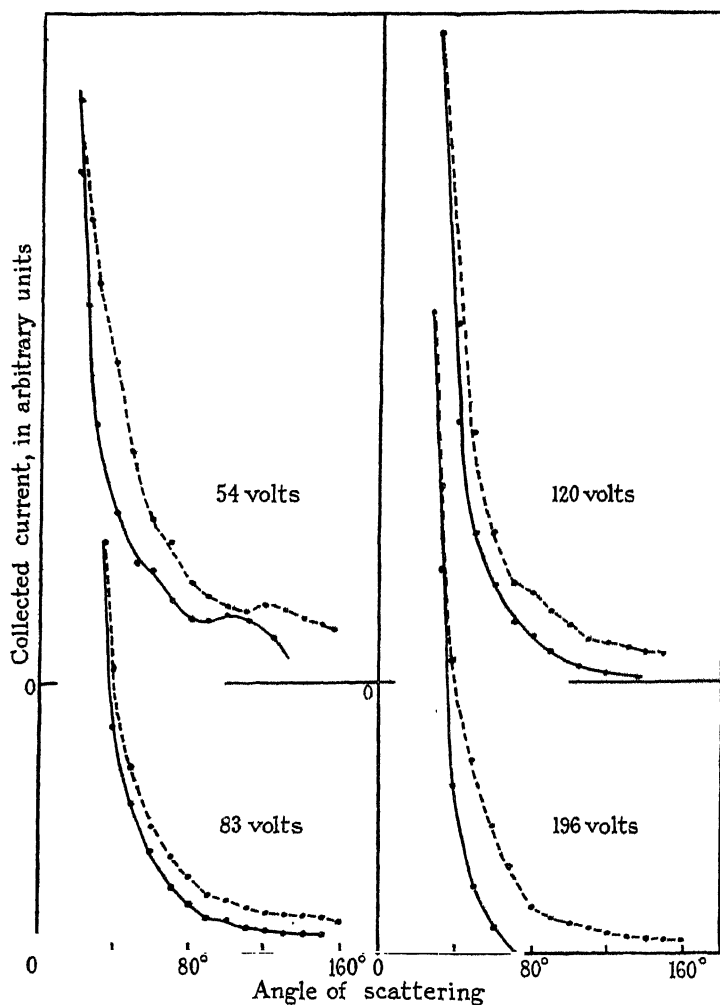


FIG. 5.—Large angle scattering in Helium.
 - - - elastic. — inelastic.

The angular distribution of the elastic and inelastic scattering (6·7 volts loss) for incident electrons of 23, 31, 42, 55, 83 and 120 volts energy is shown in fig. 7, the curves being arbitrarily fitted together at one point. It is seen

* 'Phys. Rev.,' vol. 34, p. 923 (1929).

† 'Phys. Rev.,' vol. 35, p. 1187 (1930).

‡ 'Canadian J. Res.,' vol. 3, p. 174 (1930).

that the elastic and inelastic curves closely follow one another at large angles at 120 volts and 83 volts; at 55 volts the curves are less similar, the heights of the maxima at 110° differing by a much greater amount than the experimental error. At 42 volts the curves are quite dissimilar at angles less than 80° , no maximum appearing in the inelastic curve at 60° corresponding to the maximum in the elastic curve at this angle. At 31 volts the two curves are only slightly similar, the pronounced minimum in the elastic curve at 85° appearing only slightly in the inelastic curve at 100° . At 23 volts, all maxima have disappeared from the inelastic curve, and the two curves are quite dissimilar, a

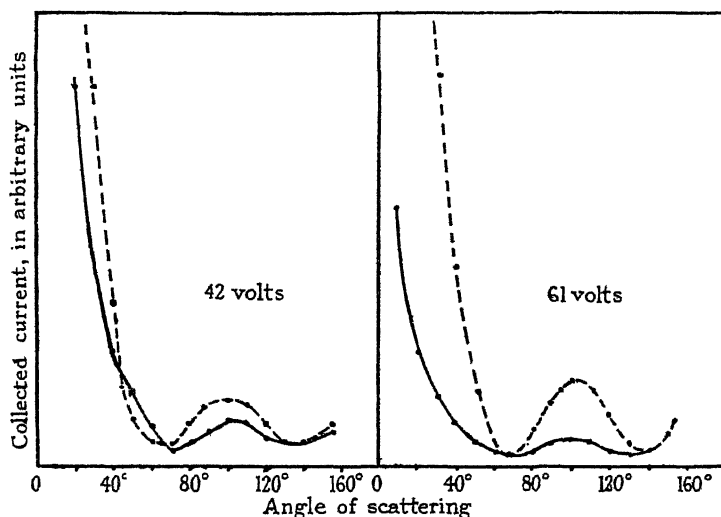


FIG. 6.—Large angle scattering in Argon.
 - - - elastic. — inelastic.

minimum appearing in the inelastic curve at the angle at which a maximum appears in the elastic curve.

The striking similarity between the elastic and inelastic curves at the higher voltages at once suggests the possibility of multiple scattering as a mechanism for the production of the similarity. Thus the results might be explained if the collected inelastically scattered electrons had suffered, not a large angle inelastic collision, but a large angle elastic collision and a small angle inelastic collision before entering the analyser.

That the results are not explicable on this basis, but are really due to a single large angle inelastic collision, seems to be clearly indicated by the following considerations.

(1) As shown in a previous section, the collected current was accurately proportional to the pressure over the range of pressures used in the experiment.

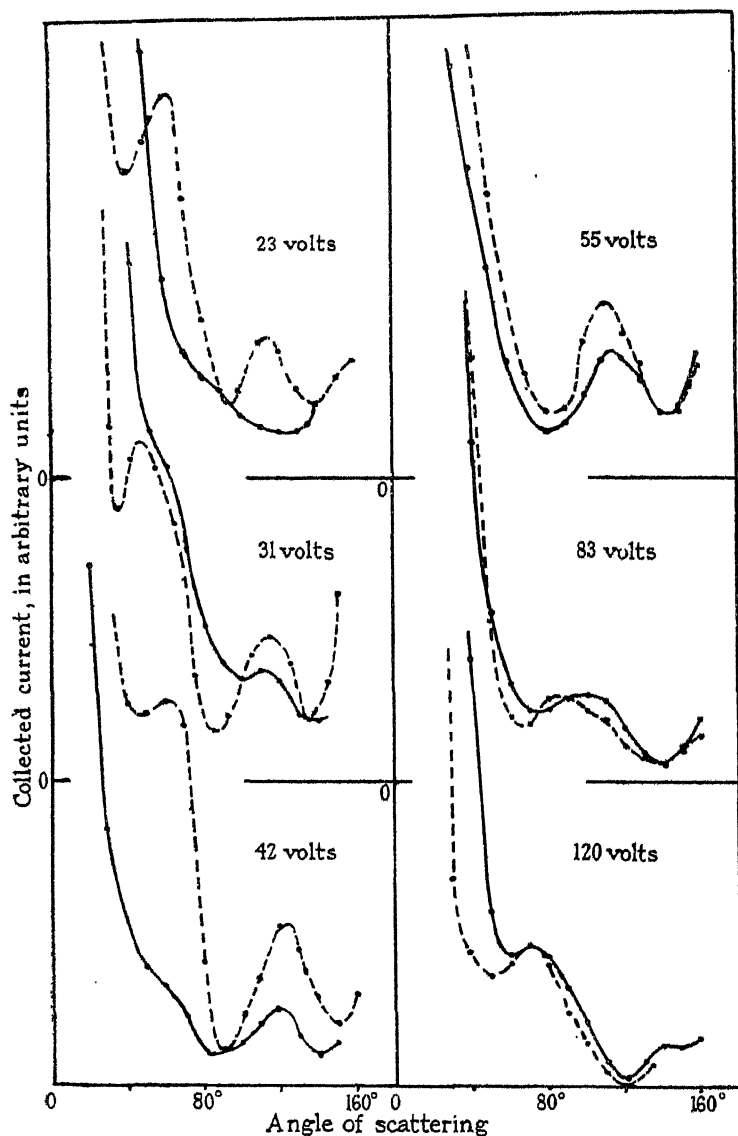


FIG. 7.—Large angle scattering in Mercury Vapour.

--- elastic. — inelastic.

(2) The intensity of the inelastic scattering at large angles in argon and mercury vapour was usually between a tenth and a twentieth part of the intensity of the elastic scattering at the same angles. If the results were due

to multiple scattering in the manner suggested above, it would require that one in every ten or twenty electrons passing across the scattering chamber should undergo a small angle inelastic collision. Calculations of the probability of such a happening on the basis of the known pressure of the gas, length of path, and the inelastic cross section, show this to be much too high a probability; actually the inelastic angular distributions in argon were found to have the form shown in the figure at a pressure as low as 2×10^{-4} mm. Hg. A further test of this point was made in the case of argon and mercury vapour by placing the gun at an angle of 0° and measuring the ratio of the height of the inelastic peak to the peak due to the main beam from the gun; in this way it was found that less than 1 per cent. of the main beam crossing the scattering chamber underwent a small angle inelastic collision.

(3) That the peaks in the inelastic angular distributions are not due in any way to the influence of the elastic angular distributions seems to be indicated by the dissimilarities which occur at the lower voltages in mercury vapour. Thus at 42 volts the inelastic curve rises very much less steeply between 80° and 30° than the elastic curve between the same angles, whereas it would be expected to rise at least as steeply if multiple scattering were responsible for the results. Again, at 23 volts there is little resemblance between the curves, a minimum appearing in the inelastic curve where the elastic curve has a peak.

Discussion of Results.

In the theory of the scattering of electrons by atoms many complicating processes are involved, and it is necessary to proceed in a series of successive approximations to include the various factors.

The first step, due to Born,* is to take as a first approximation plane waves to represent the incident electron beam; the formula so obtained is valid at sufficiently high velocities, and gives angular distributions which fall off uniformly with angle.

In fig. 4 observed angular distributions in helium are compared with those calculated using the Born formula, and it is seen that at 120 and 196 volts there is fair agreement between theory and experiment as is to be expected, for the Born formula should be nearly correct at these voltages; thus Dymond and Watson (*loc. cit.*) obtained good agreement with theory for the elastic scattering at 210 volts. It is seen, however, that the observed elastic scattering for the

* 'Z. Physik,' vol. 38, p. 803 (1926).

120-volt and 196-volt electrons is not in agreement with theory at angles as small as 20° . This is doubtless due to the fact that at these small angles, and with the dimensions of the slits used, the angular width over which the electrons are collected is comparable with the mean angle of scattering; nevertheless, the much more rapid increase of the inelastic scattering at these angles, compared with the elastic, is clearly shown.

At 84 volts and 54 volts an increasing discrepancy between the simple theory and experiment appears, the observed scattering falling more steeply with angle than given by the Born formula. Furthermore, at large angles, the experimental curves (fig. 5) are on the whole fairly flat, instead of falling to smaller and smaller values with increasing angle of scattering, as expected from the Born theory, and the failure of this theory for the heavier atoms is obvious in the region where peaks appear in the angular distribution.

It is therefore necessary to take into account the complicating effects due to the distortion of the incident and outgoing electron waves by the field of the atom, and the exchange of electrons between the incident electron beam and the atom. The effect of the distortion of the incident wave has been included in the theory by Holtsmark* and by Allis and Morse,† and good agreement with experiment obtained.

The effect of exchange on the elastic scattering has been included in the theory, together with the effect of distortion, by Massey and Mohr‡ for the case of the light elements hydrogen and helium. They have shown that if these effects are taken into account in the theory for the case of the elastic scattering, the observed effect is obtained of the curves first falling off rapidly with increasing angle of scattering and then becoming flat. The extension of the calculations§ to the case of the inelastic scattering in hydrogen and helium, taking into account the distortion of the incoming and outgoing waves and the effect of exchange, shows that a similar behaviour to that found for the elastic scattering is to be expected at the lower voltages; the calculated curves for the inelastic scattering show features similar to that exhibited by the experimental 54-volt curve, which has a maximum at 100° .

The following difficulty is now seen to arise. It has been shown‡ that it is not possible to explain the observed angular distributions in hydrogen and helium by merely including in the calculations the effect of distortion; it is

* 'Z. Physik,' vol. 55, p. 437 (1929); vol. 66, p. 49 (1930).

† 'Z. Physik,' vol. 70, p. 567 (1931).

‡ 'Proc. Roy. Soc.,' A, vol. 136, p. 289 (1932).

§ In course of publication.

necessary to include also the effect of exchange. How has it been possible, as shown by Holtsmark and by Allis and Morse (*loc. cit.*), to obtain agreement with the observed results for the elastic scattering in the case of the heavier atoms, without taking into account the effect of exchange? We shall see that the present experimental results may provide a clue.

The angular distributions calculated by such methods as the above may be expressed in terms of an infinite series of spherical harmonics, and the peaks in the observed angular distributions are due to the predominance of one or more of the spherical harmonics. Since the excited atom of mercury, say, is not very different from that of the normal atom, it seems not unreasonable to suppose that the same harmonics which predominate in the elastic scattering also predominate in the inelastic scattering at any particular voltage, provided the energy lost in exciting the atom is small compared with the energy of the incident electrons, when the wave-lengths of the incident and outgoing electron waves are nearly equal; the inelastic and elastic angular distributions would thus be similar at large angles and higher velocities. At lower velocities, however, the wave-lengths of the incident electrons and of the outgoing electrons which have lost energy will differ by a greater amount, and the similarities will tend to disappear.

Similar arguments might be applied to show that the angular distribution of the outgoing exchanged electrons would be similar to that of the outgoing non-exchanged electrons. This would probably be true down to quite low velocities, since the wave-lengths of the outgoing exchanged and non exchanged electrons are, of course, equal; thus one can see how the elastic angular distributions in the case of the heavier atoms might be explained without taking into account the effect of exchange.

The problem of the scattering of electrons by the heavier atoms may thus be simpler than would be expected from the presence of the many complicating factors, and it seems likely that in the explanation of the inelastic scattering, as well as in the elastic, it is merely the "size" of the atom which is fundamentally important.

Experiments are in progress with other gases, and also further theoretical calculations.

In conclusion, we wish to thank Lord Rutherford and Dr. J. Chadwick for their interest and encouragement throughout the work, and Dr. H. S. W. Massey for much discussion and many valuable suggestions.

Summary.

The angular distribution of inelastically scattered electrons has been investigated in helium, argon and mercury vapour for the angular range between 20° and 160° . The scattered electrons of different energies were sorted out by a uniform radial electrostatic field.

In helium, scattering curves have been obtained showing small and large angle scattering for 54, 83, 120 and 196-volt electrons, and theoretical curves are given for comparison. Angular distributions have been obtained in argon for 42 and 61-volt electrons, and in mercury vapour for 21, 31, 42, 55, 83 and 120-volt electrons. At the higher voltages, maxima and minima occur in the inelastic curves, which are similar to the corresponding elastic curves, while at the lower voltages in mercury vapour the resemblance gradually disappears.

The results suggest that the problem of the scattering of electrons by atoms is not so serious as might be expected from a consideration of the various complicating processes which are involved, and that in the explanation of the observed results for the elastic and inelastic scattering of electrons by the heavier atoms, it is merely the "size" of the atom which is of fundamental importance.

*Further Experiments on Superconductivity with Alternating
Currents of High Frequency.*

By J. C. McLENNAN, F.R.S., A. C. BURTON,* A. PITT, and J. O. WILHELM.

(Received September 12, 1932.)

In some previous papers† experiments have been described in which the phenomena of superconductivity were investigated by the use of alternating currents having frequencies up to 16 million per second. It was found that with these currents there was a critical temperature at which the high frequency resistance abruptly decreased and that this point was a little lower than the critical temperature at which superconductivity appeared when direct currents were used. This depression of the superconducting point was shown to be a function of the frequency of alternation of the currents used, becoming greater the higher the frequency. The effect was investigated with the metals tin, lead, and tantalum, by using several frequencies. In this paper further experiments upon the frequency disturbance of superconductivity are described.

1. *Experiments with Lead.*

Lead had been the metal used in the first experiments that were made with high frequency currents, and it was shown that with currents of frequency 1.1×10^7 per second there was an abrupt change of high frequency resistance as the temperature was lowered, that appeared to occur at a slightly lower critical temperature than 7.2°K. the critical point at which the direct current resistance of lead abruptly disappears. The preliminary nature of these initial experiments did not permit an accurate determination of the temperature, although there were indications that there was a depression of the critical point, such as was later established in the experiments with tin and tantalum. It was therefore desirable that the experiments with lead should be repeated and, if possible, with a more accurate determination of the temperature involved. It was thought possible that as the superconducting point with direct currents, 7.2°K. , was considerably higher than that of the other metals investigated (tin 3.76°K. , tantalum 4.3°K.) a frequency disturbance of greater magnitude might be found.

* Fellow of the National Research Council of Canada.

† McLennan, Burton, Pitt and Wilhelm, 'Phil. Mag.' vol. 12, p. 707 (1931); 'Proc. Roy. Soc.' A, vol. 136, p. 52 (1932).

The apparatus originally used (fig. 1) was modified by the addition of a helium gas thermometer at C in the space beneath the resonator, so that with it and the thermometer T temperatures noted above and below the coil might be recorded. The added thermometer consisted of a pyrex glass bulb of capacity about 10 c.c., filled at room temperature with helium at atmospheric pressure. A copper tube of fine bore sealed, at the level of the thermometer T, to a glass capillary connected the bulb to a manometer mounted outside the flask. The two thermometers were compared at three temperatures

during each experiment, namely, at room temperature, at the temperature of liquid air, and at the helium liquefaction point.

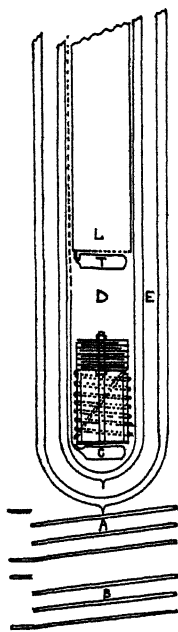


FIG. 1.

A resonator of lead wire of diameter 0.038 mm., with lead condenser plates was used, having a natural frequency 11.6×10^6 per second. The procedure followed in the experiments was exactly similar to that employed in previous work. As the resonator was cooled resonance curves were obtained at various temperatures, from the height of the "peak" of which the high frequency resistance of the lead could be deduced, being inversely proportional to the peak height. As the critical temperature was approached, the frequency of the oscillator was adjusted so that it remained at the point of resonance, and the galvanometer deflection that indicated the height of the peak was observed closely. When a sudden increase in this deflection indicated that the critical point had been reached, the temperatures instantaneously indicated by the two thermometers, above and below the coil, were recorded. Since

cooling proceeded downwards from the liquefier, above the coil, the temperature of the latter must be between these recorded temperatures, *i.e.*, the coil was warmer than the thermometer T and not warmer than the thermometer C.

The result of repeated determinations of the critical point was that the temperature of the lower thermometer was $7.0 \pm 0.1^\circ$ K. while that of the upper was $7.0 \pm 0.1^\circ$ K. The indicated point was then 7.0° K. with a possible error of $\pm 0.1^\circ$ K. Repeated experiments with direct currents had shown that the change to superconductivity occurred when the upper thermometer registered a temperature of 7.2° K., *i.e.*, when the temperature of the specimen was certainly not less than this. There was then a definite

indication of a depression with this frequency of about 0.2° K. The difficulties of regulating and of reading accurately the temperature of the circulating gas prevent more definite measurement of the effect.

2. *Bismuth-Lead Alloy.*

An alloy consisting of 58 per cent. bismuth and 42 per cent. lead had previously been found to have a superconducting point that was above 8° K.* The transition to superconductivity was known to be very abrupt, the ratio R/R_0 just before the critical point being as great as 0.5, dropping to a zero value in a fraction of a degree Kelvin. This feature made it particularly suitable for measurements with high frequency currents, where it had been found that the higher the resistance of the resonator, the more definite the indication of the critical point became.

Some wire of this composition was therefore made by an extrusion process by Baker & Co., Newark, N.J. A fibre former was wound with it to make a resonator of natural frequency corresponding to the frequency 11.6×10^6 . For condenser plates tin was used, as being more easily worked and available than plates of the alloy. As chief interest was in the temperature at which the transition to superconductivity took place rather than in the values of R/R_0 for the alloy, this could introduce no relevant error. The arrangement of the two thermometers was the same as that used in the experiments with lead just described.

The indication of the critical point proved to be very definitely given by the sudden increase in height and sharpness of the resonance peaks. Fig. 2 shows, in the contrast between the resonance curves taken at 10° K. and at 4.2° K., how very different was the effect upon the oscillator when the metal was in the superconducting state from that when it was just above the critical point. The results with the frequency of 11.6×10^6 per second were that the upper thermometer registered a temperature of 8.1° K. and the lower thermometer 8.3° K. at the critical point. With conditions as nearly as possible identical, the experiment was repeated using direct currents. For this measurement potentiometer and current leads had to be added to the coil at the junctions of the wire with the condenser plates. The corresponding temperatures were found to be practically identical with those obtained with the high frequency currents, namely, 8.1° K. for the upper thermometer and 8.35° K. for the lower. Though there was a slight indication, in the last reading, of a depres-

* McLennan, Allen and Wilhem, 'Trans. Roy. Soc. Canada,' Sec. III, vol. 24 (1930).

sion of the superconducting point with frequency, it was not sufficiently great to be measurable by the methods available at these temperatures.

The frequency disturbance of superconductivity, in the metals studied, was seen therefore to be an effect of small magnitude, measurable with accuracy only at temperatures below the boiling point of helium, where the more sensitive measurement and control of temperature by vapour pressure could be employed.

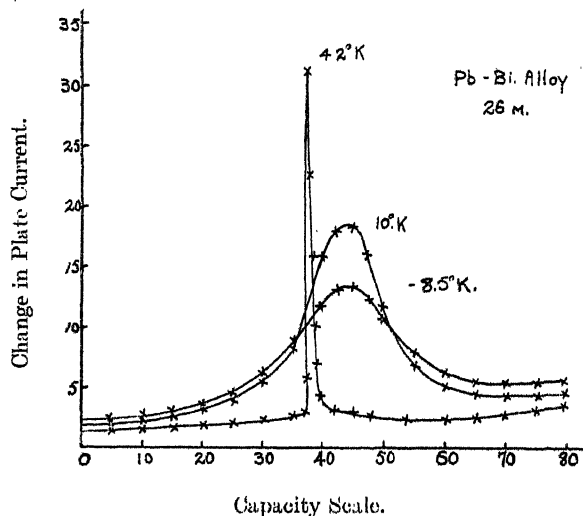


FIG. 2.

3. *Experiments with Higher Frequencies.*

An experiment was made to find the depression of the superconducting point of tantalum with currents of a higher frequency than any yet used, namely, 30.3×10^6 per second corresponding to a wave-length of about 10 metres. The natural frequency of the resonator of tantalum used with oscillations of the lower frequency was increased to this frequency by removing some of its condenser plates. It was found necessary in the case of the generator, however, to change the circuit to that known as the Hartley type. The variation in frequency was made by a variable condenser connected across the end of the coil that was in the "plate circuit." The frequencies were measured on a wave meter which was calibrated directly by measurements of wave-length on a Lecher wire system.

No difficulty was found in observing with accuracy the temperature at which the resonance peaks by increasing abruptly, indicated an abrupt decrease in the high frequency resistance of the metal. With direct currents this abrupt

decrease occurred at a vapour pressure of 920 mm. of mercury, *i.e.*, at a temperature of 4.4° K. well above the boiling point of helium ; but with currents of the high frequency used the critical point was not reached until the pressure was reduced to 770 mm. of mercury. The corresponding points at which the transition to superconductivity was complete were for the direct currents, 907 mm. and for the high frequency 650 mm. There was therefore a frequency depression of the critical temperature of about 0.2° K., equivalent to a difference in vapour pressure of the liquid helium of over 150 mm. of mercury. The results so far obtained for tantalum are tabulated below.

Frequency.	Starting point.		Finishing point.	
	Pressure.	Temperature.	Pressure.	Temperature.
0 (direct currents)	920	4.41	907	4.39
11.6×10^6	820	4.28	705	4.12
30.3×10^6	770	4.21	650	4.04

It will be seen that there is an indication that the critical temperature-frequency curve follows a course similar to that found for tin. If this be so, and the curve between the frequency 11.6×10^6 and 30.3×10^6 be taken as linear, extrapolation would give as with tin, a frequency of the order of 10^9 per second as the one corresponding to a critical temperature of 0° K. An experiment with a still higher frequency than 30.3×10^6 per second is, of course, necessary to establish this point beyond any doubt.

Experiments with Direct and Alternating Currents flowing simultaneously in the Specimen.

Since the experiments had shown that, when alternating currents of high frequency were used, the superconducting state was not established until a lower temperature was attained than that reached when direct currents were used, it was therefore of interest to inquire whether the addition of high frequency currents would prevent the direct current resistance also from disappearing until the lower temperature was reached. On the other hand, the presence of the direct current might cause the metal to become superconducting to the high frequency currents also at the normal, direct current, critical point, or some intermediate result might be found.

Since the resonator did not offer a closed circuit to direct currents, there was no difficulty in the measurement of the direct current resistance of the coil.

For this purpose two pairs of leads, one pair to carry current to the coil, the other for the measurement of the drop of potential across the coil, were attached to the ends of the latter where they made connection to the condenser plates. This addition to the high frequency circuit, however, effectively "short-circuited" the condensers of the resonator and destroyed the possibility of obtaining resonance in that circuit with the generated oscillations. Chokes were therefore to be inserted in all four of the added leads to prevent the passage

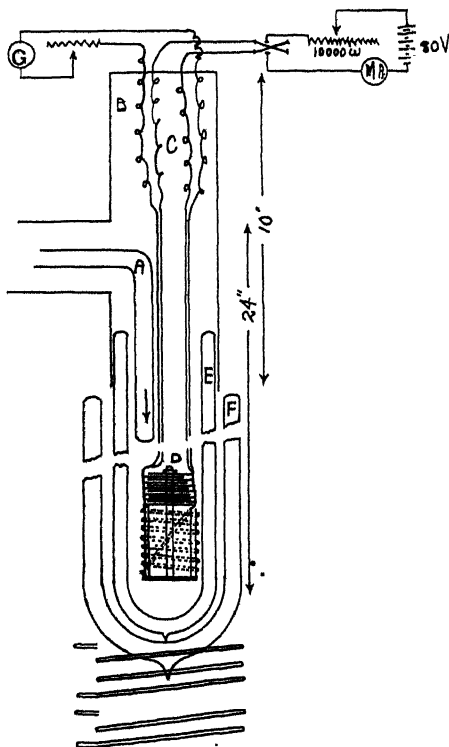


FIG. 3.

of the high frequency currents by that path. After considerable experimentation an arrangement was achieved in which the two measurements, that of the direct current resistance by the potential drop across the coil, and that of the high frequency resistance by the observation of resonance curves, were very nearly independent of each other at room temperature.

The arrangement finally adopted is shown in fig. 3. The resonator was contained in the side flask into which liquid helium could be siphoned over by the tube A from the flask in which it was liquefied. The chokes C were contained in the metal head B of the side flask. They were made by "space

winding" constantan wire of No. 40 gauge upon asbestos string, a length of choke being then fixed by thread to a fibre card. This was found to be a convenient method of obtaining sufficient impedance with small distributed capacity in a compact form. The resistance of each of the chokes was about 200 ohms. To avoid disturbances, due to thermal electromotive forces, the leads D to the resonator were of the same constantan wire. Although for the blocking of high frequency currents it would be desirable to place the chokes down in the flask closer to the resonator, this could not be done as the heat generated in them by the direct currents used in an experiment would boil away the helium. The direct currents were driven by a battery of cells through a high resistance of the order of 10,000 ohms, included in order that any changes in the resistance of the chokes as the temperature was lowered might have negligible effect in altering the currents in the specimen. For the measurements of the potential drop across the coil, the observation of the deflection of a high resistance galvanometer shunted by the specimen was preferred to the usual potentiometer measurement as it was of greater simplicity. Errors due to thermal electromotive forces were eliminated, by observing the deflections in opposite directions on the scale of the galvanometer when the reversing switch, K, was operated. The deflections of the two measuring galvanometers, the one in the direct current circuit and the other in the plate circuit of the generator, were shown on two scales that were arranged one immediately above the other.

Tantalum was chosen as the metal to be used for the resonator because of its purity and of the position of its critical point.

When the tantalum coil had been covered with liquid helium the pressure upon the liquid was raised to a total greater than 950 mm. of mercury; the normal transition temperature for tantalum being that corresponding to 920 mm. vapour pressure. With the high frequency generator not operating, observations of the deflection of the galvanometer giving the potential difference across the tantalum coil were taken at successively lower pressures until superconductivity was complete. The graph, fig. 4, between the resistance ratio R/R_0 and the temperature shows the results in the curve marked by circles. The change in resistance began to be evident at a pressure of 927 mm. and was complete at about 907 mm. pressure. The direct current in the specimen was 12.9 milliamperes.

The high frequency generator was then switched into operation, its frequency being adjusted until the peak of the resonance curve indicated that the maximum high frequency current was being induced in the tantalum coil. While

the maximum high frequency current was flowing observations of the direct current resistance were then repeated over the same range of pressures as above. No change from the previous results could be noticed. In order that the ratio between the induced high frequency currents and the direct current in the metal might be made as great as possible, the coupling between the generator and resonator was increased while, at the same time, a smaller direct current, 7.1 milliamperes, was used.

The curve marked by crosses in the graph, fig. 4, shows the results now obtained. It is evident that the presence of the high frequency currents

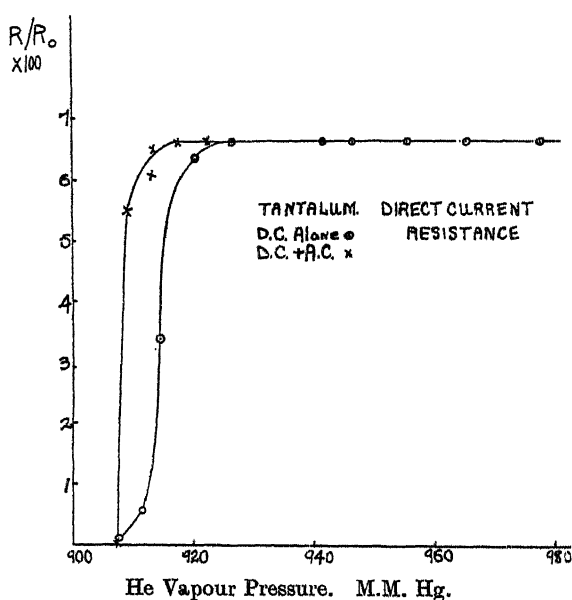


FIG. 4.

delayed the initial appearance of the change in resistance, although the final point where the resistance had reached a zero value was unaltered. The greater "steepness" of the transition when the high frequency currents were present was very noticeable when the experimental observations were being made.

This result is what would be expected if there were an effect of the high frequency currents in preventing the change to the superconducting state until a lower temperature than the normal. Due to the "skin effect" the high frequency currents would be largely confined to the outside of the tantalum wire, leaving a central region of the wire where the high frequency current

was very small, however great the total high frequency current carried by the wire might be. We were then measuring the resistance of a number of conductors in parallel, in which there were high frequency currents of varying magnitude from zero up to the maximum current density at the surface of the wire. There was therefore one of these conductors, a fine core in the centre of the wire, whose resistance would follow the curve marked by circles, vanishing at the normal point, at 907 mm. pressure. The measured direct current resistance, namely, that of the conductors in parallel, had therefore to vanish at this point. At higher pressures where the resistance of the filament considered had not completely vanished, the increased resistance of the other filaments, due to the high frequency currents in them, made the total measured resistance ratio R/R_0 higher than that found without high frequency currents present. The curve obtained was therefore just what would be expected of a system of conductors in parallel under such conditions.

To complete this part of the experiment, an estimate was made of the relative magnitudes of the high frequency and direct currents present in the wire during the experiment. The flasks around the resonator and chokes being removed, a small "crossed-wire" type of vacuum thermocouple of resistance about 0.6 ohm was inserted in series into the circuit of the resonator at the point where one end of the tantalum coil was connected to a set of condenser plates. The leads from the junction were short lengths of fine wire, leading to a galvanometer and twisted together to minimise induction of high frequency currents in them. The maximum deflection of this galvanometer when the generator was operated at the resonance point was noted, and the magnitude of the direct current was found that would produce the same deflection when passed through the resonator coil. This direct current was 17.5 milliamperes. Since it was known that for the fine wire of the thermocouple the high frequency resistance was negligibly greater than the direct current resistance, the high frequency currents must have been also of this magnitude, 17.5 milliamperes. The resonance peak recorded by the variation of the plate current of the generator was simultaneously of about half the value obtained in the actual experiment at the low temperature. Since, with constant coupling, between resonator and generator the height of the resonance peaks was proportional to the current in the resonator (by the relation $e = M \frac{di}{dt} = -M\omega i$), the value 35 milliamperes can be given as the lower limit of the high frequency currents in the resonator at the low temperatures. The direct current was only one-fifth of this in the final experiment.

This point is made in some detail because it gives an estimate of the order of magnitude of the high frequency currents used in the whole series of experiments that have been made with the different metals, in all of which the currents, as shown by "resonance peaks" had values less than the 35 milliamperes just arrived at. The possibility of the depressions found in these experiments being due to errors, caused by heating effects that do not occur in direct current experiments, is therefore denied beyond doubt. For in the experiment just described, high frequency currents of magnitude greater than any previously used were present yet produced no apparent depression of the superconducting point at all comparable to that measured in the experiments referred to.

It was logical to proceed to an experiment in which no part of the metal under test could escape the action of the high frequency currents. Skin effect may be practically eliminated by the use of a tubular conductor in which the thickness of the metal is small compared to the diameter. Such a conductor was constructed by "wiping" a layer of block tin upon a constantan wire, of diameter 0.016 cm. The tin skin was of average thickness about $1/500$ mm., and its presence decreased the resistance of the wire at room temperature by about 7 per cent. Calculation shows that at the low temperatures just above the superconducting point the resistance of the constantan would then be at least 30 times that of the tin, so that the presence of the core of constantan could introduce no serious errors in the measurements in that region of temperature. This device had the advantage also that the resistance of the resonator at the low temperatures was much greater than it had been possible to obtain by the use of solid tin wire.

The experimental procedure was the same as in the preceding experiment. The deflection of the galvanometer indicating the potential drop, proportional to the direct current resistance of the coil, was observed at a series of temperatures with and without the presence of high frequency currents simultaneously flowing in the metal. The results are shown by the graph, fig. 5, in which the curves show the relation between the direct current resistance ratio R/R_0 and the temperature, both without high frequency currents present, and with these simultaneously flowing in the metal.

When, in addition to the direct current, high frequency currents were induced in the same conductor, the resistance was changed so that the curve AB first obtained was shifted to lower temperatures, becoming the curve A'B'. The switching on and off of the high frequency generator changed the resistance reversibly from the point A to A', B to B' and so on. Even when the resistance had become zero on the undisturbed curve, it could be partially restored to the

metal by the switching on of the generator. The points on the curves were obtained as the temperature was gradually lowered by decreasing the vapour pressure of the helium, but in addition a series of pairs of points such as AA', BB', were obtained as the specimen was allowed to warm up. With these the accuracy of the curve could be checked.

The ratio of the high frequency to the direct current was then decreased by simultaneous decrease of the coupling of generator to resonator and increase of the direct current in the specimen; curves intermediate between the two just described were obtained showing that the position of the displaced curve depended on the relative magnitudes of the direct and high frequency currents in the specimen. A number of curves were obtained with different values of this ratio, but for clearness only one of them, CD, is included in the graph.

High Frequency Resistance.

In the same experiment it was possible to follow the changes in the high frequency resistance quite independently of the measurements of the direct current resistance. When observations giving the latter were recorded at any temperature the height of the resonance peak, indicated by the second galvanometer, were simultaneously observed. Curves can, therefore, be drawn showing how the high frequency resistance changed with temperature, both with and without accompanying direct currents in the specimen. Two of such curves are drawn in fig. 6, curve I being that obtained without, curve II with, accompanying direct current in the metal. The high frequency resistance can be given in arbitrary units only, as the total high frequency resistance, which included an "induced" resistance, due to the neighbouring currents as well as the resistance of the tin-constantan coil, was the quantity that was inversely proportional to the height of the resonance peaks. The relative values of the resistance of the coil itself plotted in the graph were obtained on the assumption that the added "induced" resistance was a constant quantity in this range of temperature.

It was seen that the presence of the direct current had the effect of shifting the curve to higher temperatures, so that while there was still some resistance offered to the high frequency currents alone, that resistance could be wholly or partially removed by the addition of the direct current field in the metal, as shown by complete change A to B and partial change C to D, fig. 6. When both currents were flowing in the metal, the respective temperatures at which the high frequency began to change abruptly and had completed the change,

were the same as the corresponding temperatures for the direct current resistance that was simultaneously observed in the experiment. This could be verified by inspection of the values obtained in the various experiments with different ratio of high frequency to direct current. That is, observations of the direct current resistance taken at the same time as the observations giving the high frequency resistances in curve II, fig. 6, yielded values of the direct current resistance ratio R/R_0 that lay on a curve intermediate between curve AB and A'B' in fig. 5 (close to, but not identical with the curve CD).

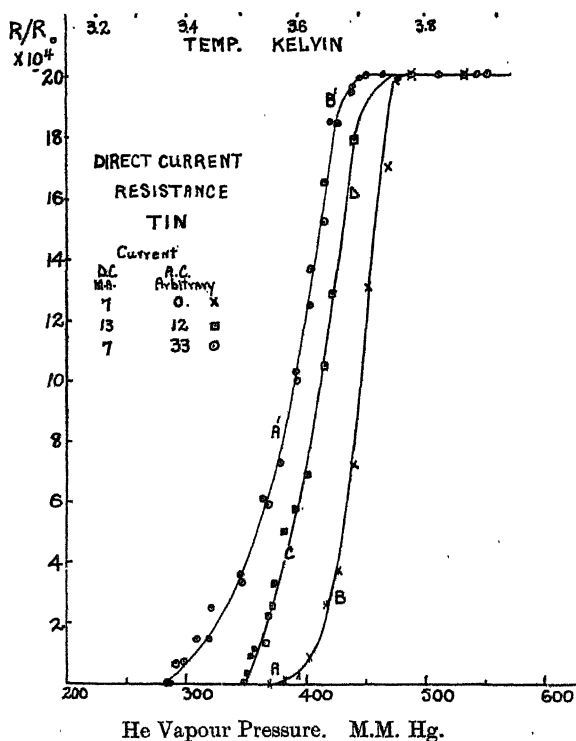


FIG. 5.

Two effects have therefore been established, the depression of the critical point for the direct current resistance by the application of high frequency currents, and the raising of the critical point for the high frequency resistance in the presence of a direct current. When both currents are flowing in the metal, the critical point for the direct current resistance is the same as the critical point for the high frequency resistance, i.e., when the superconducting state had been established at this common critical point, resistance was offered neither to the direct nor to the alternating current. The position of

this common critical point on the temperature scale was dependent on the ratio of the magnitude of the direct to the high frequency currents in the specimen.

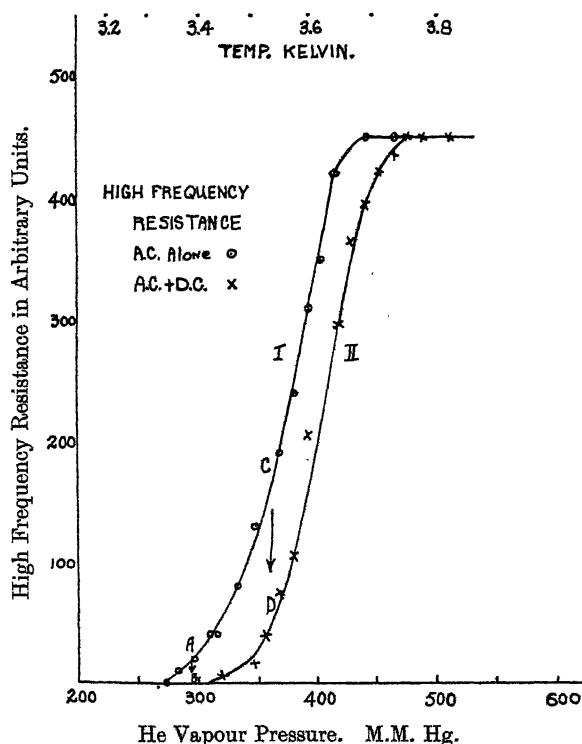


FIG. 6.

Summary.

Experiments have been carried out on the phenomena of superconductivity with alternating currents of high frequency, following those already reported. In these experiments, observations have been made on the resistance of a conductor at low temperatures when both alternating (frequency 12×10^6) and direct currents were flowing simultaneously.

The experiments may be divided into two sets :

- (a) The resistance offered to direct currents by the metal (tin) was measured both with and without accompanying high frequency currents.
- (b) The resistance offered by the same sample to high frequency currents was measured both with and without accompanying direct currents.

The resistances of the conductor to direct currents and to high frequency currents were measured by independent methods.

When both currents were flowing, the critical point for the high frequency resistance was the same as the critical point for the direct current resistance ; the position of this common critical point on the temperature scale is determined by the ratio of the magnitude of the direct to that of the alternating current. Thus, when the superconducting state has been established at this common critical temperature, resistance was offered neither to direct nor to alternating currents.

Two effects have therefore been established, the depression of the critical point for the direct current resistance by the application of high frequency currents, and the raising of the critical point for the high frequency resistance in the presence of a direct current.

These experiments confirm the reality of the frequency disturbance of the superconducting point found in our early experiments, and it follows that any theory of the nature of superconductivity that may be advanced must include an explanation of this new phenomenon.

Adsorption. A Study of Availability and Accessibility.

By Sir WILLIAM HARDY, F.R.S., and MILLICENT NOTTAGE.

(Received March 16, 1932.)

Our object is twofold—to show how friction may be used to analyse the composition of an adsorbed layer, and to exhibit the remarkable effect of “previous history” upon its composition and properties. When polar molecules are absent the composition of the layer when first formed commits it to a path of change from which it can escape only by turning back on its path and passing through the point of origin.

In the opinion of one of us the experiments are of interest to biologists. They reveal systems, let us call them solutions, whose relations to an adsorbing surface are complex and active only over a curiously narrow range of temperature of some 15 to 20 degrees; outside this range, temperature has no effect. The influence of “previous history” is great and that of chemical constitution obvious. The remarkable “specific” relations so characteristic of living matter are not there but only molecules of simple type incapable of mutual chemical reaction are involved:—Specificity in fact is merely a question of degree. Given molecules of complex atomic pattern and the phenomena would probably have been quite beyond analysis, indeed the curves for the normal reaction (adhesion) of simple binary mixtures are beyond explanation when the components can react together or when one is a ring and the other a chain compound.*

The facts will be stated in terms of two assumptions. It will be assumed that the slider used, which had a spherical face, owing to the great pressure under it, reduced the layer of lubricant to two primary layers with the yield plane or surface of slip between them. This is the first assumption; its validity has been discussed in earlier papers.† Static friction, which alone is considered, is the strength of the joint in shear; it is also the maximal reaction of the surface of slip to traction.

The second assumption is that each of the molecules which compose the two primary layers contributes independently to the total reaction, and the value of an individual contribution is fixed by the atomic structure of the molecule to which it belongs. This assumption makes it possible to identify the species

* Nottage, ‘Proc. Roy. Soc.,’ A, vol. 118, p. 607 (1928), figs. 2 and 3.

† Hardy and Bircumshaw, ‘Proc. Roy. Soc.,’ A, vol. 108, p. 2 (1925).

of molecule present in the layers and to follow changes in the composition of the latter.

The equation for the friction of a lubricant composed wholly of a single long chain compound is :

$$\mu = \frac{b' + b''}{2} - [d + c(N - m)],$$

where the first term on the right is a pure function of the nature of the solids, steel, glass, etc., and the term in brackets is a pure function of the atomic structure of the lubricant.* Since the solid was the same throughout, both slider and plate being of mild steel, the contribution of the solid can be neglected in computing changes in the relative numbers of different species of molecules.

A few words must be devoted to the first assumption. In 1912 one of us† pointed out that in all close packed structures whose molecules are other than spherical the cohesive forces across an interface would not only hold molecules by direct attraction, the range of the force being of the order of dimensions of a molecule, but would also penetrate as an oriented strain transmitted from molecule to molecule.‡

Direct observation with X-rays and indirect evidence such as is furnished by *e.g.*, the Mean Value Rule which defines the effect of the nature of the solid upon friction and adhesion, have confirmed this view, at any rate for long chain compounds and for solid surfaces.

The layer held by direct attraction is probably identical with the primary layer—that is the well-known layer of insensible thickness which spreads over clean fluid or solid faces. This layer has a large heat of formation, is strongly held and is probably mono-molecular. It is the layer formed by “primary spreading.”§

But the mechanical properties of the layers formed by transmitted strain are also determined mainly by the fields of force of the solids. Indeed the Mean Value Rule shows that the influence of a solid can penetrate some thousandths of a millimetre into a fluid in contact with it. Therefore as one has primary and secondary spreading of liquids over surfaces, it is just to distinguish primary adsorption and secondary adsorption.

According to the first assumption, however, the lubricating system is reduced by a spherical slider to two primary layers, one on each steel face, with the

* Hardy and Bircumshaw, *loc. cit.*, p. 24.

† Hardy, ‘Proc. Roy. Soc.’ A, vol. 86, p. 131 (1912).

‡ Hardy and Doubleday, ‘Proc. Roy. Soc.’ A, vol. 100, p. 568 (1922).

§ Hardy, ‘Phil. Mag.’ vol. 38, p. 49 (1919).

surface of slip between them. Only steady states are considered, and as the slider was frequently moved about on the plate the primary layers may be considered to have been in equilibrium with the oil.

In the sequel it will be seen that both assumptions are possibly too narrow, for even a spherical slider seems sometimes to be unable to penetrate to monomolecular layers, and the mechanical properties of a molecule in the primary layer depends not only upon its own chemical constitution and the solid to which it is attached but also upon both normal and tangential reactions with its neighbours.

Certain terms need definition. By successive extractions of two commercial oils with acetone or by exposure to adsorbing surfaces such as those of clean glass beads it has been found possible to remove the more reactive constituents until the fraction dissolved by the acetone or adsorbed had the same composition as the residual oil.* Since selective adsorption had then ceased the residuum was an *adsorption individual* with no temperature coefficient of friction between the limits explored, namely, 10° C. to 106° C. From its method of preparation it may be taken to have been freed from polar molecules.

The adsorption individual was found to form from 93 to 96 per cent. of the two commercial oils examined—it is the *diluent* which holds in solution from 4 to 7 per cent. of active *components*.†

Obviously when a lubricant is a single pure chemical individual it is also an adsorption individual and has no temperature coefficient of friction. No exception to this rule has been found.

A temperature-friction curve is said to be *reversible* when it has the same form for an ascending as it has for a descending series of temperatures. The curves for adsorption individuals are always reversible.

The *state* of an oil is defined by the statistical average taken by volume or by time of the molecular species composing it. The word in short has its ordinary significance in kinetic theory. The state of a component in an oil may be such that in competition with other components or with the diluent itself it cannot find a place in the adsorbed layers—it is then said not to be *available* for adsorption. It may, however, be available when it competes for a place on a *clean* surface but unable to displace other molecules already in possession of the surface. It is then said not to be *accessible*.‡

* Hardy and Nottage, "Lubrication Research, Technical Paper No. 1," p. 16, H.M. Stationery Office (1929).

† *Ibid.*, p. 39.

‡ Grammatically monstrous but serviceable.

Availability and accessibility are two potentials which are not absolute but purely relative to the similar potentials of the diluent, and therefore if they could be measured on some absolute scale a variation in, say, the availability of the component might be found to be due either to a change in its own state, or to a change in the state of the diluent. But as the experiments do not allow us to discriminate, the simpler course has been taken of stating the facts wholly in terms of the component as though the variations were absolute and not merely relative.

A convenient diluent for experimental purposes is the medicinal "paraffin" of the pharmacopœia which is here called "Oil B.P." It is prepared from a mineral oil by prolonged percolation through highly adsorptive material. It has no temperature coefficient of friction over the range explored, 8° to 106° C. It appears to be composed of saturated cyclic compounds and has the great advantage of combining a low vapour pressure with fluidity over the whole range.

Besides oil B.P. a wax denoted by N.O.P. was used as diluent. It also has no temperature coefficient of friction, and appears to be a mixture of saturated hydrocarbons melting between 54° and 57° C.

The respective coefficients of friction are, for the mild steel used :—

$$\mu_{\text{B.P.}} = 0.228 \quad \mu_{\text{N.O.P.}} = 0.275.$$

The components used were palmitic acid $\mu_{\text{ac.}} = 0$, melting point 62° C., hexa-decyl alcohol $\mu_{\text{al.}} = 0.114$, melting point 50° C., and the wax N.O.P.

The interest of the experiments lies in the comparison of two entirely distinct adsorption equilibria. The first is the equilibrium when the oil in mass, the plate and the slider were at the same temperature before the pool of oil was placed on the plate. It gives the adsorption for a particular temperature *when the adsorbing surfaces are clean*. The friction-temperature curve will for short be called the X curve and be shown as an interrupted line.

The second is the equilibrium when the system has been warmed or cooled to the particular temperature whilst the oil is in contact with the adsorbing surfaces. The curve is called the O curve and is drawn as a continuous line. Obviously the O curve always starts from that point on the X curve at which the oil has been applied to the surfaces.

A point on the X curve gives the value reached when lubricant and adsorbing surfaces have been brought to that temperature separately and maintained at that temperature long enough to be in equilibrium with it.

A point on the O curve on the contrary gives a value reached when the adsorbing surfaces have been continuously in contact with the oil and in equilibrium with it through a range of temperature. It is a value determined

not merely by the availability of the substances for adsorption but also by their accessibility, that is by their capacity for displacing molecules already in possession of the surface.

The X curve is a curve of availability whose shape depends upon the state of the oil in mass, the O curve is in the main a curve of accessibility.

The X curve was reversible in each of the three examples studied, that is to say the friction at any temperature was the same whether the mixture in bulk had been warmed up, or cooled down to that temperature. Curve O was reversible only when no polar molecules were present.

Both curves are remarkable for two things, the variations of friction lie within a narrow range of temperature and are limited to the friction of the pure diluent on the one hand and of the pure component on the other.* In discussing the former care has been taken not to hide ignorance by the use of the word colloidal.

Example 1.—No Polar Molecules. Diluent, Oil B.P. Component, Wax N.O.P.

The wax was dissolved with vigorous stirring in the oil at a little above the melting point of the former. The mixture was then allowed to cool and kept for 24 hours. To the naked eye it seemed homogeneous. Three mixtures were studied containing respectively 2.4, 0.51, and 0.16 per cent. of wax.

(i) *Wax 2.4 per cent. (Fig. 1.)*

Curve X.—Since the curves for oil B.P. and wax N.O.P. are horizontal lines at $\mu = 0.228$ and $\mu = 0.275$ respectively, the figure shows that curve X leaves the former at about 22° and joins the latter at about 45° C. Therefore below 22° the wax in solution is not available for adsorption whilst above 45° the oil B.P. is not available.

Between these limits and as temperature rises the state of the mixture must change in a way which increases the availability of the wax so that more and more molecules find their way into the primary layers. The form of the curve, indeed, suggests that availability is limited to some particular species of molecules of the wax and that the concentration of this particular species increases with rise of temperature.

The change with temperature of the relative number of molecules of oil B.P. and of wax in the primary layer might be due in part to a change in the selective adsorption of the steel surface. This is unlikely because it seems

* The limitation holds only for steady states, see Appendix, p. 281.

incompatible with the fact that an adsorption individual prepared from a complex oil by adsorption at a particular temperature or by extraction with acetone has no temperature coefficient of friction over the full range. A single chemical individual also has no temperature coefficient.

The changes in state of the oil were completely reversible, that is to say, the position of a point in the curve was the same when the oil was warmed from

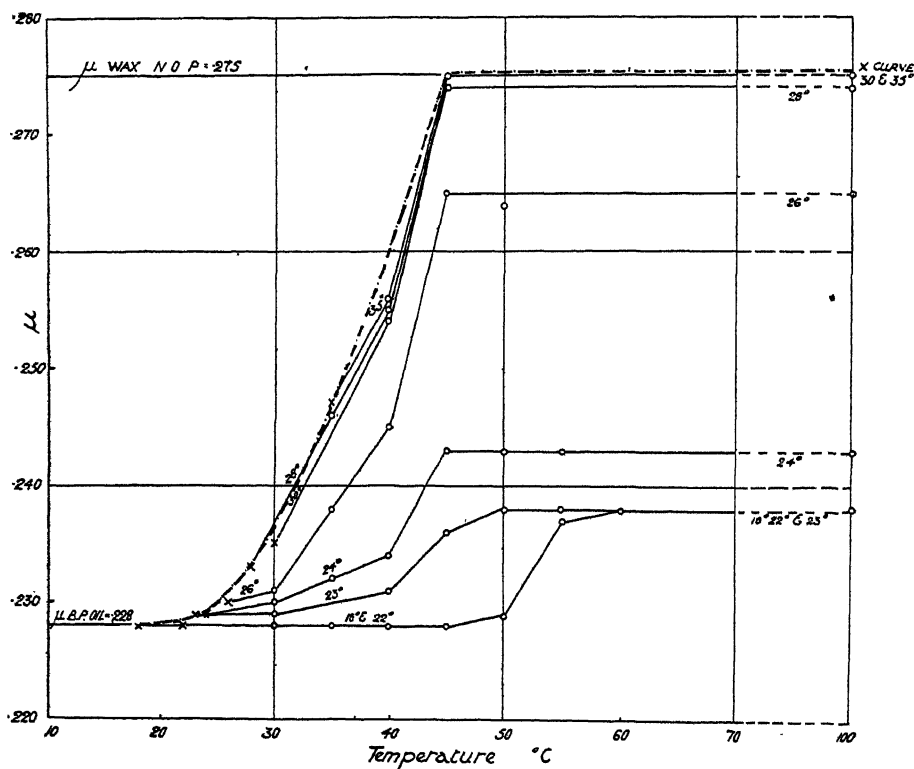


FIG. 1.—In this and in the succeeding figures, the X curve is an interrupted line; the O curves are in continuous line with the initial temperature given in numerals.

room temperature as it was when the oil was cooled from 100° C. In this special sense then curve X was reversible.

Curve O must always start from some point on curve X, and its form, as fig. 1 shows, depends upon this initial temperature which, as curve X shows, fixes the initial composition of the primary layers. The form of curve O therefore depends upon the capacity of molecules of wax to displace molecules of diluent.

It may be assumed that this capacity increases as availability increases, but if this were all curve O would coincide with curve X. Since it does not so

coincide its form must depend upon the variation with temperature both of availability and accessibility.

The fundamental assumption enables us to give a numerical value to the availability of the component which may be defined as the fraction of unit area of a primary layer seized by the component when it competes with the diluent for position on a clean solid surface. The X curve then shows that availability is zero at 22° C. and has unit value at 45°. At intermediate temperatures the value is given by the ratio $\mu_o - \mu_{B.P.} \div \mu_{N.O.P.} - \mu_{B.P.}$.

Turning to fig. 1 it will be seen that when the initial temperature is at or below 22° and the primary layers are composed wholly of molecules of oil B.P. the molecules of wax are unable to get access below 48° C. in spite of the fact that availability is unity at 45°.

When the initial temperature is about 30° or higher the O curve coincides with the X curve, accessibility is therefore complete and the curve follows the change in availability.

Between these two limits there are a set of curves which have remarkable features. Each sweeps up to a saturation level at which the composition of the primary layer becomes independent of a further rise of temperature and the critical temperature at which this condition is reached is sensibly 45° when availability has unit value.

It will be noticed that the curves sweep up more steeply and the saturation level rises rapidly as the initial temperature rises. From this we must conclude that the presence of molecules of wax in the primary layer when it is first formed increases the chance of access of further molecules of wax. It is as though each molecule of wax weakened the hold of the molecules of B.P. oil in its neighbourhood.

From the fundamental assumption the fraction of unit area of a primary layer occupied by wax molecules can be calculated. The values are given in columns 3 and 5 of Table I.

The O curves are completely reversible and the equilibrium therefore must be kinetic—that is to say it is the result of continuous evaporation from and condensation on to the primary layer from the overlying oil. The reversibility has certain striking features, which can be indicated briefly by considering the curves at the limits of the series. To save space a particular curve is indicated by a subscript which gives the initial temperature—that is the point on curve X at which curve O starts.

Curve O_{18° to 22°}.—The curve has the same form when plotted by an ascending series of temperatures as it has if the series is 18° → 100° and then by

Table I.

Curve X.

No availability.
Below 23° $\mu = \cdot 228$.

Availability unity.
45° $\mu = \cdot 275$.

Curve O.

Initial temperature.	Accessibility begins.	$\frac{\mu_{\theta} - \mu_{B.P.}}{\mu_{N.O.P.} - \mu_{B.P.}}$	Temperature saturation level.	$\frac{\mu_{sat.} - \mu_{B.P.}}{\mu_{N.O.P.} - \mu_{B.P.}}$
18° 22°	48° $\mu = 0\cdot228$	0·00	55° to 60° $\mu = 0\cdot238$	0·21
23°	23° 0·229	0·02	47° 0·238	
24°	24° 0·229		45° 0·243	0·32
26°	26° 0·230	0·04	45° 0·264	0·77
28°	28° 0·233	0·10	45° 0·274	0·98
30°	30° 0·235	0·15	45° 0·274	1·0
35°	35° 0·247	0·19	45° 0·275	

descending steps to 18°. Its form is fixed by the starting-point only. The kinetic character of the equilibrium is seen when a big jump in temperature is made, there is then a latent period of about 20 minutes during which the friction rises or falls, as the case may be, to the steady value.

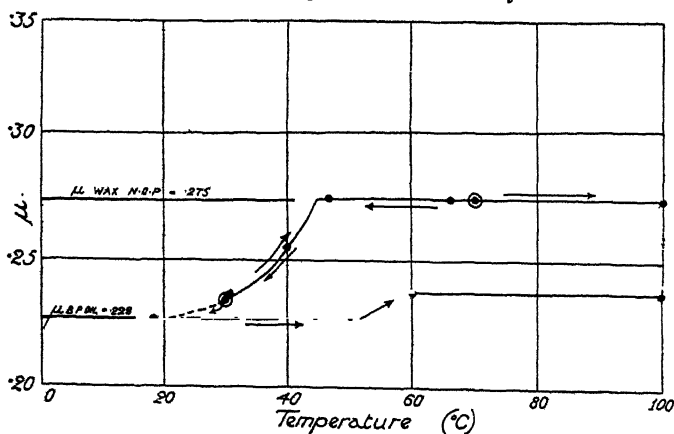


FIG. 2.

Curve O_{30°}.—So long as the series starts at 30° the points lie on the same curve independent of the order in which they are taken, but if the lower limit of the curve is passed the system takes on an entirely new character: fig. 2 will make this clear. Two curves are drawn in continuous line, the curve O_{30°}

for ascending temperatures and the curve O_{18° . To avoid confusing the figure all the points observed are not noted—they may be taken to be those shown in fig. 1.

A mixture containing 2.4 per cent. wax was applied to the clean steel faces at 30° and the friction measured at temperatures up to 100° and down to 18° and the points, as will be seen, lie on the O_{30° curve. At 18° , however, all the wax molecules have left the primary layers and the system now is the same as though the initial temperature was 18° , for on warming again the points lie wholly on that curve.

The initial temperature obviously commits the system to a path of change from which it seems able to escape only by overstepping the lower limit of the path.

It seems to be impossible to put the system on any of the paths O_{22° to O_{30° save at the lower limit, but the path of O_{30° can be entered at any point simply because availability and accessibility are both at limiting values. For this reason the O curve was found to be the same for initial temperatures 30° , 35° and 70° .

One important conclusion remains to be noted: it is that at any temperature above that at which the availability of the component begins (22° C.) an indefinite number of equilibrium values of friction are possible for any one temperature, each depending only on the previous history of the system.

The form of the curves invites many questions which cannot be answered without more knowledge. Why has the initial state such a preponderating influence? If, as one is almost driven to admit, the molecules of wax initially in the adsorbed layer are centres of increased accessibility, why does the process of replacement not go on as it were autocatalytically until the limit is reached when the primary layer is wholly composed of wax? Saturation levels are relatively easily understood when the component is composed of polar molecules and the O curves are completely irreversible, but how a path of change fixed only by an initial state can coexist with complete reversibility is difficult to explain. On the simplest possible view a saturation level would be due to condensation of all the wax into the adsorption layer, but that layer must then be more than mono-molecular, since a rough calculation, in which the smear of lubricant was taken at its thinnest, shows that such a layer could contain no more than 4 per cent. of the wax. An arithmetical explanation of all the facts could no doubt be framed in terms of the ratio between wax dissolved and adsorbed but that would miss the real interest, at any rate to biologists, which is the organisation of the whole of the oil (compare pp. 269, 279 and 281).

(ii) Wax 0.51 per cent. (Fig. 3.)

Curve X is less steep and the saturation level much lower, availability rising only to about 0.7. The temperature limits are, however, the same, namely, 23° to 45° C. The O curves show the same general features as with 2.4 per cent. wax, but the levels are lower.

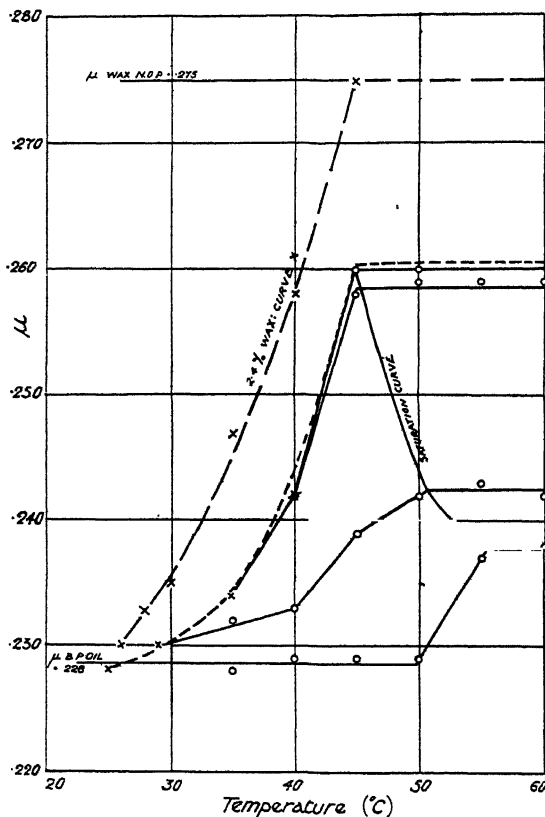


FIG. 3.—0.51 per cent. wax. The X curve for 2.4 per cent. wax is shown for reference.

Availability is maximal at 45° C. but, owing to the small quantity of wax, the concentration of the “active” molecular species is not high enough to displace completely oil B.P. from the primary layer.

(iii) Wax 0.16 per cent. (Fig. 4.)

The saturation levels are still lower, that of the X curve having fallen to $\mu = 0.242$ when the availability is only 0.3. The temperature limits have changed, the curves being moved up the scale.

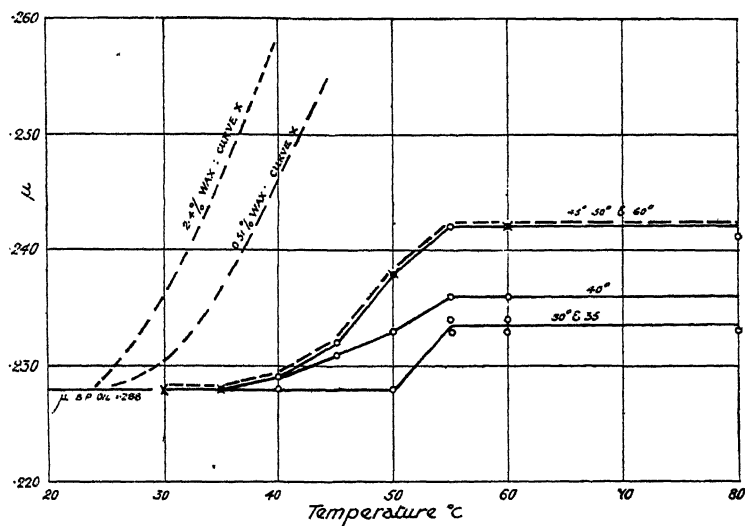


FIG. 4.—0.16 per cent. wax. The beginnings of the X curves for 2.4 per cent. and 0.51 per cent. are shown for reference.

Comparing the three concentrations we get :—

X Curve.

Wax.	Temperature limits.	Accessibility complete.	Friction range.
per cent.	°	°	μ
2.4	23–45	30	0.228–0.275
0.51	23–45	40	0.228–0.260
0.16	40–55	45	0.228–0.242

The shifting of the curves to the right by a fall in total percentage of the wax raises new difficulties. Since solubility increases with rise of temperature the maximal dispersion of the wax would not be moved up the temperature scale by decreasing the total percentage, therefore a rise of temperature must do something more than merely break down complexes of molecules.

The effective concentration of wax depends to only a small extent upon the total concentration. This at any rate is the conclusion to be drawn from a detailed study of curve O_{18} .

The percentage of wax was varied between the limits 0.13 and 3.06. The form of the curve was the same for all the mixtures and the curves practically

coincided from about 0.5 per cent. upwards. Below 0.5 per cent. the saturation level falls and the temperature at which the primary layer is saturated with wax falls as the total concentration falls, but the change is slight. The temperature at which accessibility begins is sensibly independent of concentration.

Table II.—Initial Temperature 18° C.

Wax.	Accessibility begins.	Saturation.	Saturation level.
per cent.	°	°	μ
3.06	50	60	0.239
2.40			0.239
1.47			0.239
0.52			0.238
0.25		55	0.236
0.13		55	0.234

Enough points to fix the temperatures exactly were not measured.

The temperature of access and indeed the whole group of curves are related closely to the melting point of the wax. The curve O_{18} was plotted for two waxes and two normal paraffins and fig. 5 shows the close relation in each case between availability, accessibility, and the melting point.*

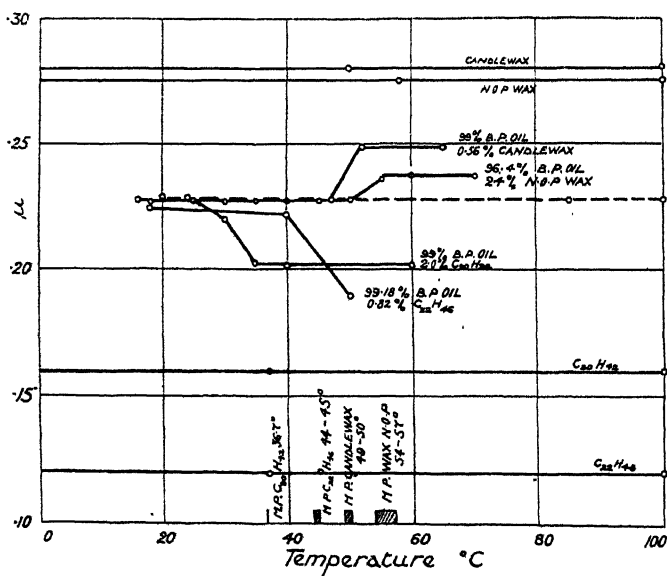


FIG. 5.

* When the report ("Lubrication Research, Technical Paper No. 1") already referred to was written the only curve known to us was curve O_{18} and the relation to the melting point led to the inference that it was a simple curve of availability. Curve X and the other curves show that the inference was wrong.

The broad conclusion is that the limits within which the composition of the adsorbed layer becomes a rapidly varying function of temperature are fixed mainly by the appearance in the mixture of highly dispersed molecules of the component, but this is not the only way in which temperature intervenes. The "active" species begins effectively to compete with the diluent at 22° C.

Example 2.—Component, highly-polar, namely, Palmitic Acid. Diluent, non-polar Oil B.P. Melting point 63°. $\mu_{ac.} = 0$, $\mu_{B.P.} = 0.228$. (Fig. 6.)

At first sight nothing could be more different than the behaviour of this system. The curves in example 1 sweep upwards, in 2 downwards, the former are wholly reversible, the latter are wholly irreversible and yet there are fundamental similarities. In both it is obvious that the limit within which the curve can move is fixed by the friction of the two constituents. In both

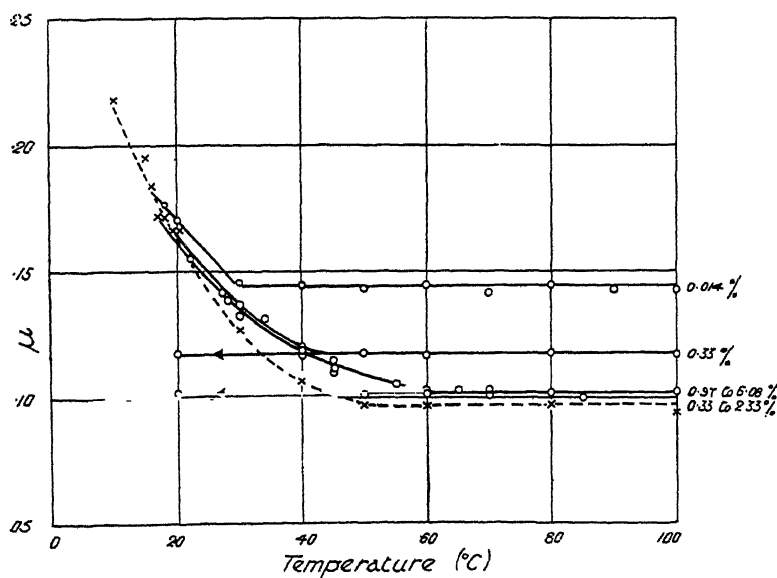


FIG. 6.

there are saturation levels, but the most remarkable and unexpected feature of this polar system is the fact that it is not possible, at any rate within the limits of concentration tried, to do what can be so easily done with the non-polar system, namely, to reach the limit at which the friction is that of the pure component. This is exactly what would not be expected from *a priori* considerations because it is the characteristic habit of polar molecules to seize the whole of the adsorbed layer.

The mixtures were made as before by dissolving the acid in the oil at a little above the melting point of the former. The apparently homogeneous solutions were then kept at room temperature for 24 hours. From 0.97 per cent. upwards some crystals of acid separated out at 18°.

The possible limits for the curves are $\mu_{B.P.} = 0.228$ and $\mu_{ac.} = 0$, but the latter, as fig. 6 shows, is never reached.

Curve X.—Availability begins at about 9° and is greatest at about 50° or rather less, the observations not being close enough to fix the point exactly. The form of the curve is the same for mixtures containing from 0.33 to 2.33 per cent. of acid. It is therefore not sensibly affected by the composition of the mixture. The melting point of palmitic acid is 62°, at which point oil and acid seem to be miscible in all proportions, availability is, however, greatest below this point. The curve shows that whilst availability can be zero at $\mu_{B.P.} = 0.228$, it cannot be pushed to unity at $\mu = 0$. Instead a saturation limit is reached at $\mu = 0.097$, which is independent not only of a rise of temperature but also of the percentage of acid in the mixture. This curious limitation was not imposed by the particular diluent for it appeared when the wax N.O.P. was so used. The values were :—

Diluent.	Per cent. acid.	Saturation level.
B.P. oil	0.97 to 6.08	= 0.097
Wax N.O.P.	0.33 to	= 0.090
	2.2	= 0.089

Curve X is completely reversible in the special sense mentioned on p. 263. It is so important to realise what this special sense is that it will be well to put it into quite other words. In curve X the steel face as an adsorbing surface is used to measure the state of the mixture in bulk and the variations of state with temperature were found to be reversible.

Curve O.—It was unfortunately difficult to hold temperatures below room temperature with the apparatus, for this reason the effect of varying the initial temperature was not determined. The effect of varying the concentration of the wax was, however, followed somewhat closely for the curve O_{18} . Mixtures containing the following percentages by weight of acid were used :—0.014, 0.33, 0.97, 1.5, 2.41 and 6.08. In order to make the figure legible the readings for 1.5 per cent. and 2.41 per cent. are omitted. The general form of the curve is the same for mixtures containing from 0.014 to 6.08 per cent. acid and the curves coincide from 0.97 to 6.08 per cent.

Accessibility is slightly less than availability throughout. The saturation level varies widely with changes in the composition of the mixtures up to 0.97 per cent. and then becomes independent of composition. Between the limits 0.0 and 0.97 per cent. there is an infinite number of curves, the saturation level falling and the saturation temperature rising with increase in the percentage of wax. The curves were completely irreversible, that is to say, once the saturation level was reached, the curve for falling temperature was a horizontal line.

When we recollect that mixtures containing less than 0.97 per cent. seemed homogeneous to the naked eye throughout the range of temperature, it seems to follow that availability is limited to a particular species of molecule, the concentration of which increases as the temperature rises. The most obvious suggestion again is that this molecular species is the completely dispersed unassociated molecule and that the concentration of the available or "active" species increases with rise of temperature.

A saturation level stopping short of complete replacement is indeed an anomaly when strongly polar molecules are concerned. For example, the isotherm of a mixture of caprylic acid dissolved in undecane is given in fig. 7

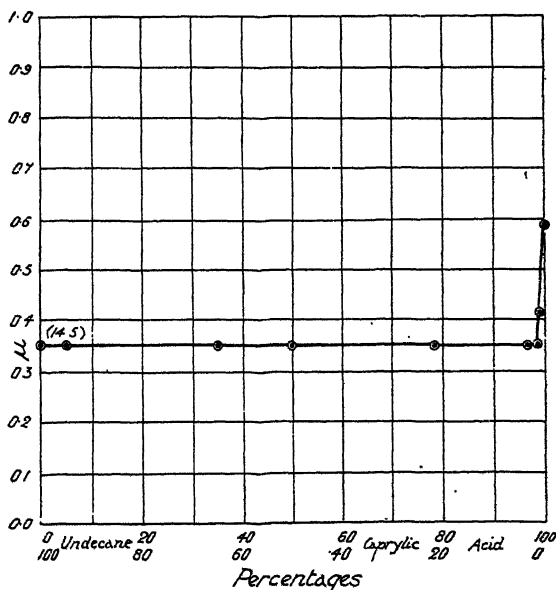


FIG. 7.—Showing the effect of the relative concentration of acid and paraffin upon friction.

and it shows that the non-polar molecules do not obtain access to the dissolved layer until the percentage of acid has fallen to less than about 0.7 per cent.

For concentrations higher than this the friction is the same as that of the pure acid. The same kind of curve is given by mixtures of paraffins and alcohols. In fig. 8 the isotherms for solutions of palmitic acid in oil B.P. are reproduced from Technical Paper No. 1 of the Lubrication Committee, and it is remarkable that each has the general form of an adsorption curve, but with the unusual feature that the saturation limit for the acid never reaches 100 per cent. and is a function of temperature.

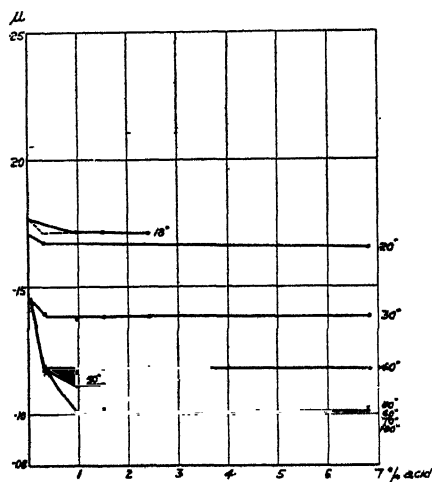


FIG. 8.—From "Lubrication Research, Technical Paper No. 1," by courtesy of the Controller, H.M. Stationery Office.

The hypothesis that availability is determined by the degree of dispersion of the acid opens up an interesting possibility. Availability reaches its limit somewhat short of the melting point of palmitic acid when dispersion is likely to be maximal. The limit is, however, not unity, that is to say, the acid fails completely to displace the diluent in the primary layer. It looks as though the highly-polar molecules acted in the opposite way to the non-polar molecules of wax by increasing the hold on the steel face of the molecules of diluent in their neighbourhood. This is intelligible if each polar molecule acts as a dipole which induces polarity in its neighbours. It will be remembered that Sutherland based his electrical theory of cohesion in liquids on such induced polarity.

Availability may, however, be fixed by secondary adsorption. It is obvious from mere inspection that solutions of polar molecules in oil B.P. when in contact with the steel are very different from the mixture in bulk. The oil ceases to be uniform, it readily breaks up into greasy-looking patches. When the component is an alcohol there is the same gross interference with the state

of the mixture. No such gross disorganisation of the oil, however, was noticed in the absence of polar molecules.

Example 3.— $\mu_{al.} = 0.114$. Component polar, namely Cetyl Alcohol. Melting point $50^{\circ} C$. Diluent, Oil B.P. $\mu_{B.P.} = 0.228$. (Fig. 9.)

Two mixtures only were used containing respectively 0.39 and 2.20 per cent. of the alcohol. The limit of the temperature range explored was 18°

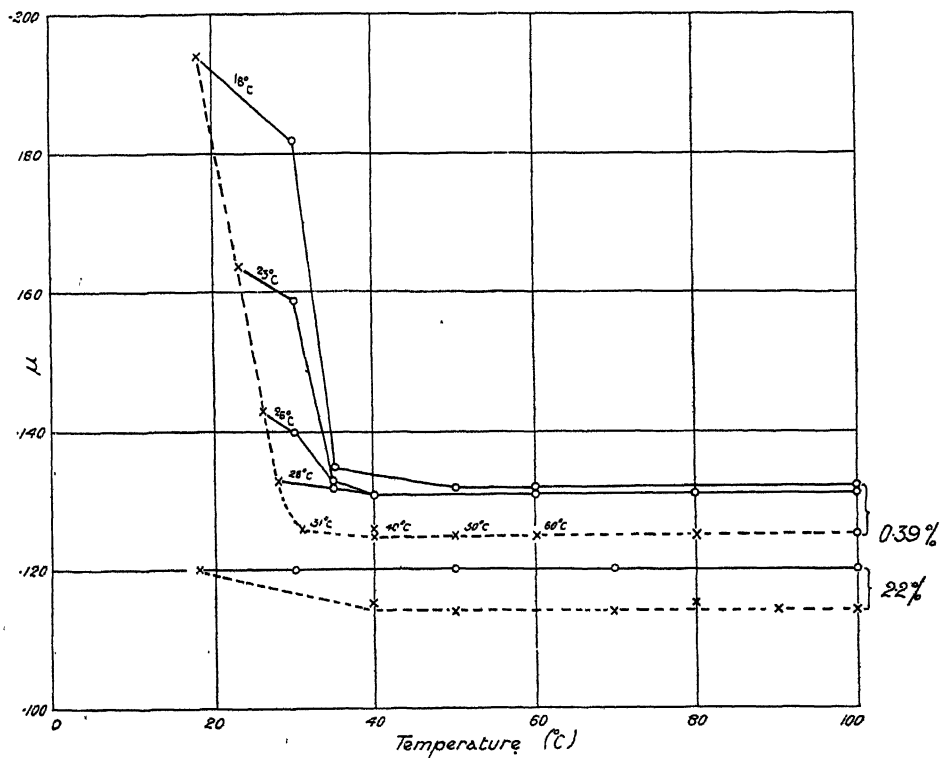


FIG. 9.—The numerals indicate initial temperature. The O curves from 31° upwards coincide with each other and with the X curve. To avoid confusion, the individual measurements are not marked.

to 100° . The 2.20 per cent. mixture was rather jelly-like at room temperature (18° to 20°) and on standing showed some tendency to deposit crystals and become more fluid.

When put on the plate at any temperature below about 60° a visible thick film formed which dulled the highly polished surface of the steel. On this film droplets of liquid floated. The film was there above 60° , as the presence of droplets showed, but it was transparent.

When a thin layer of solid lubricant, which is also a single chemical individual,

is deposited on the plate from solution in a volatile solvent it is necessary to break away the upper crystalline crust by moving the spherical slider about slowly in order to get the steady normal* readings of the primary film.† The way in which the slider had to be moved about to get steady readings gave the impression that the film mentioned above had sufficient thickness and tenacity to hinder it from penetrating to the primary layer.

Since the friction of the pure component was below that of the diluent the curves sweep downwards. All the curves were irreversible.

X Curve.—This curve was plotted between the limits 18° – 100° for both mixtures. Availability reached unity with the 2.2 per cent. mixture, but failed to do so with the 0.39 per cent. mixture, the saturation level being at $\mu = 0.125$.

O Curve. 0.39 per cent. Mixture.—Within narrow limits the saturation level depends upon the initial temperature and therefore upon the initial proportion of alcohol in the primary layer. Accessibility is maximal at about 30° , the curves $O_{30^{\circ}}$ and $O_{40^{\circ}}$ coinciding with the X curve.

2.2 per cent. Mixture.—The curve $O_{18^{\circ}}$ is singular in that it lies throughout at a saturation level. Accessibility is independent of temperature and the increase in availability to unity at about 30° is without effect.

These curves again raise the old difficulty. If the spherical slider really reduces the lubricant to two mono-molecular layers they must have contained about 6 per cent. of diluent at 18° when the total percentage of alcohol was 2.2. But since the O curve is horizontal the rise of availability in the oil in bulk to unity at 30° did not dislodge the diluent, possibly because continuous contact with the steel face so changed the whole pool of oil as to prevent availability reaching unity.

Lowering the total percentage of alcohol to 0.39 shifts the curves to the right, but if the maximum of availability were due only to a maximum degree of dispersion one would expect it to be reached with the lowered concentration of alcohol either at the same or even at a lower temperature; therefore a rise in temperature must act in some other way. It is, of course, possible that the slider fails to penetrate the tenacious film sufficiently to reach a mono-molecular layer, and the difference between the curves for the 0.39 and the 2.2 per cent. mixtures may be due to the fact that they refer to surfaces at different levels.

* "Normal" because they are independent of further movement of the slider, and because they fall on the curve connecting molecular weight and friction.

† Hardy and Doubleday, 'Proc. Roy. Soc.,' A, vol. 100, p. 559 (1922); Hardy and Bircumshaw, *loc. cit.*, p. 20.

Example 4.—Components, Palmitic Acid and Wax N.O.P., Diluent, Oil B.P. (Fig. 10.)

Systems with two active components do not really come within the limits set to this enquiry, but one such is worth mentioning because it confirms the distinction between availability and accessibility. Wax N.O.P. was added to a mixture of oil B.P. and palmitic acid containing 0.34 per cent. of the latter. The only effect was to lower the friction on the X curve, leaving both the saturation temperature (40°) and saturation level of the O curve ($\mu_{\text{sat.}} = 0.118$) unchanged, fig. 10. The availability of the acid at the initial temperature (18° to 20°) was therefore increased until when 9 per cent. wax was present it was maximal, *i.e.*, it had the saturation value.

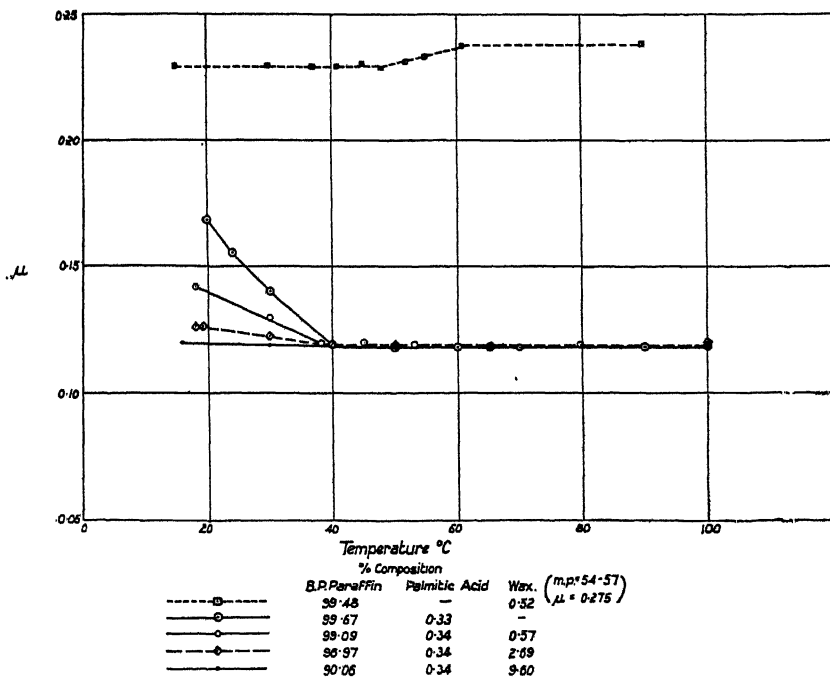


FIG. 10.—From "Lubrication Research, Technical Paper No. 1," by courtesy of the Controller, H.M. Stationery Office.

Wax itself, however, did not enter the adsorbed layers at all and the curves were completely irreversible.

Palmitic acid is soluble in paraffin wax and the presence of the latter increases the solubility of the acid in B.P. oil. There can therefore be little doubt that it increased the availability of the acid by increasing the grade of dispersion at the lower temperatures but without altering the saturation level.

From these examples of simple adsorption we may conclude :—

(1) That only in a highly dispersed state can a component compete with the diluent for a place in the primary layers, and the limits within which adsorption varies rapidly with temperature are fixed mainly by the appearance in the oil of a particular molecular species of the component, possibly a simple chemical molecule, and by the rise in concentration of this species until it obtains complete possession of the primary layer or is maximal for the total concentration of component.

(2) But temperature must intervene in another, as yet unknown, way.

(3) The simple two-dimensional hypothesis which considers only a monomolecular adsorbed layer in equilibrium with liquid in mass is not enough, the conception must be three-dimensional and take account of the obvious gross re-organisation of the whole pool of oil which contact with the steel face brings about. This re-organisation must be due to polarisation which spreads from molecule to molecule along the normal from the steel face, it is the kind of polarisation which one of us has called “diachysis.”* It may also be necessary to admit a polarisation which spreads from molecule to molecule tangentially.

(4) The first assumption (p. 259) may be too narrow. Even a spherical slider may sometimes fail to penetrate to monomolecular layers.

Greasy Patches.—When either of the mixtures which contained polar molecules was spread on the plate patches of higher viscosity formed. The higher the temperature and the greater the percentage of the polar component the sooner did the patches appear. Their formation was hastened by moving the slider about—that is to say by stirring the mixture on the plate.

With hexa-decyl alcohol the patches tended to take the form of lenses on a continuous film which below about 60° was thick enough and opaque enough to dull the burnished surface of the steel. Above 60° the film was not visible, but it was there as the presence of the patches proved.

The phenomenon is without doubt related to the instability of a layer of immiscible liquid spread upon water.

When the slider was moved through a grease patch it left a track or groove owing presumably to the high viscosity.

No non-polar mixture has been found to form grease patches.

Latent Period.—The duration of the latent period, that is the time taken by the oil to get into equilibrium with the steel was shortened by moving the

* Hardy, ‘Phil. Trans.’ A, vol. 230, p. 30 (1932).

slider about. The time was undoubtedly related to the gross changes in the oil which developed the patches. It is important to note that mechanical agitation shortened the latent period of non-polar mixtures in just the same way therefore, though there was no formation of obvious grease patches there undoubtedly was a re-organisation of oil in the pool.

The latent period of non-polar mixtures was quite as long as that of polar mixtures and in this connection it is to be remembered that while neither wax N.O.P. nor oil B.P. alone had any latent period the polar bodies, palmitic acid and hexa-decyl alcohol taken alone have each a latent period of orientation of 60 and 40 minutes respectively.

The latent period, the grease patches and the rise in viscosity taken together do, in our opinion, prove that contact with the adsorbing surface produced molecular structure throughout the whole pool of oil. It is for this reason that the hypothesis of a monomolecular adsorbed layer in equilibrium with oil in mass is too narrow, secondary adsorption must be reckoned with.

On the O curves the latent period of the non-polar mixture for rising temperature was one of rising friction and the opposite for falling temperature ; it lasted about 20 minutes.

With the two polar systems the latent period for rising temperature was always one of falling friction and lasted 20 to 50 minutes depending upon concentration and temperature. But for falling temperature there was no latent period because there was no change in the composition of the primary layer. It will be remembered that once a saturation limit was reached the friction became constant over the whole range of temperature.

Movement of the slider was strikingly different with polar and non-polar mixtures. With non-polar mixtures it was jerky, the joint was brittle the slider breaking away suddenly and chattering forward. The property called "oiliness" was absent at any rate for mild steel. Movement with the polar mixtures was a smooth glide.

Owing to the brittle nature of the joint it was difficult to get readings with the non-polar mixtures. They were very sensitive to vibrations.

PART II.—THE FUNDAMENTAL ASSUMPTION.

The effect of applying a lubricant to a solid face is to reduce the friction of the latter. This is readily perceived when two solid faces are coated separately with *solid* lubricant and then brought together. This "external" friction has the same value as the internal friction of a liquid joint when the same

lubricant is used at a temperature at which it is solid and at a temperature at which it is liquid.

The equation quoted on p. 260 shows that the observed value of static friction, that is of the greatest tangential reaction to traction, of which the surface of slip is capable, depends upon the nature of the solids. Adhesion, that is the greatest normal reaction to a surface of break to tension, also depends upon the nature of the solids and it can be proved that the solids do not contribute directly but by polarising the molecules of lubricant.*

Similar proof for the tangential reaction is not forthcoming but it can safely be assumed that what is true of the normal is true also for the tangential reaction. It is assumed therefore that the tangential reaction is the sum of reactions between individual molecules of lubricant acting across the surface of slip.

This assumption offers the simplest explanation of the following facts.

When the lubricating film is deposited from vapour the tangential reaction is a linear function of the density of the vapour until the latter is saturated when the value is the lowest possible.† This seems to prove that the adsorbed film in equilibrium with saturated vapour is a close packed structure.

A single chemical substance has no temperature coefficient of friction, the close packed structure of the primary layer being independent of temperature.

No adsorption individual has been met which has a temperature coefficient of friction.

The form of the X curve which lies between limits defined by the friction given by the pure constituents of the mixture, shows a gradual replacement of one kind of molecule by another in the adsorbed layers as temperature changes.

It follows that throughout a plateau in a curve the primary layer must have the same composition, whilst ascending or descending parts of a curve show that its composition is changing.

Summary.

(1) The variation of static friction with temperature, of three solutions in medicinal paraffin was measured, the solutes being a wax, palmitic acid and hexa-decyl alcohol.

(2) Variation was limited to a curiously narrow range of temperature outside of which friction was invariant. The limits of variation were the friction of the pure solvent, called in the text the "diluent," and that of the pure solute,

* "Lubrication Research," *loc. cit.*, p. 29.

† Hardy and Doubleday, *loc. cit.*, p. 556.

friction therefore could be used roughly to measure the proportion of solvent and solute molecules in the adsorbed layer.

(3) For a given temperature the friction and therefore the composition of the adsorbed layers are determined by previous history and the composition of the solution.

(4) Two changes govern all the phenomena—change in state of the bulk solution which fixes the *availability* of the solute for adsorption and a change in the capacity of the solute for replacing adsorbed solvent, which is the *accessibility* of the solute.

APPENDIX.

Note on the Latent Period of Lubrication.

The latent period during which the lubricant is coming into equilibrium with the attraction fields of the solid is important in practical boundary lubrication, indeed there are some indications that the greater part of the wear of bearings occurs in it.

The phenomena of the latent period are of the simplest when the lubricant is a single pure long chain compound. Variations of friction which precede the steady state may be due either to variations in the thickness of the layer of lubricant or to orientation of its molecules. The former can be eliminated by using a spherical slider which cuts through at once to its final level. When this is done the duration of the latent period is, in minutes :—

Normal paraffins	0
Normal alcohols	} about 40
Carbinols	
Normal acids	„ 60

and friction always falls as orientation completes the adsorbed layer. .

Adhesion moves in the opposite way to friction, it increases as the steady state is reached.

The only ring compounds examined, however, namely, phenanthrene and naphthalene gave a latent period of about 40 minutes with falling adhesion. This is at one with a curiously opposed behaviour as between ring and chain compounds noted in an earlier paper on friction.*

These simple relations recall, and are akin to the simple behaviour of chemical individuals on a surface of water, glass or metal. All surface relations become excessively complicated, however, when a mixture is used unless it happens to be an adsorption individual. The spreading of a complex oil on water is an

* 'Phil. Mag.' 6, vol. 40, p. 201 (1920).

example. There is not only the lag due to the slowness of tangential diffusion as was first pointed out by Willard Gibbs but there is also the slow exchange by evaporation and condensation between the first formed adsorption layers and the body of the mixture.

The mixtures dealt with in the foregoing paper are simple but the phenomena of the latent period are characteristically complex. There is a latent period of orientation for both non-polar as well as polar mixtures.

When the oil is first put on the plate the friction is abnormally high being outside the limits always observed in the steady state. The friction therefore falls at first and the rate of fall is increased by moving the slider about. The effect of such mechanical disturbance has been shown by X-ray analysis to hasten the orientation of long chain molecules in layers with the chain more or less at right angles to the solid surface.* During the latent period the readings are not only high but irregular and the slider breaks away much more jerkily than it does in the steady state. The following figures are quoted merely by way of illustration.

Temperature 24°. Steel on steel, slider spherical. Lubricant 2·3 per cent. paraffin wax N.O.P. dissolved in oil B.P.

Time.	Interval.	μ
h. m.		
4 36	0	—
4 47	11	0·305
5 60	24	0·291
5 9	33	0·239
5 19	43	0·260
5 30	54	0·247
5 41	65	0·261
5 49	73	0·226
5 58	82	0·229
6 3	87	0·229
6 9	94	0·230
6 14	99	0·228

} Latent period

} 0·228

Temperature of chamber raised to 50° by 6 hours 25 minutes.

Time.	μ
h. m.	
7 1	0·244
7 12	(0·239)
7 24	0·244
7 30	0·243
7 42	0·244

} 0·244

* Bragg, 'Proc. R. Inst..' vol. 24, p. 481 (1925).

The low reading in brackets was due to accidental vibration. The limits within which the curve for steady states varies are $\mu_{B.P.} = 0.228$ and $\mu_{N.O.P.} = 0.275$, therefore the first two readings in the latent period, which occupied about 70 minutes were above them, and such abnormal readings were commonly obtained.

On Plasticity and Creep in Solids.

By HAROLD JEFFREYS, F.R.S., Reader in Geophysics, Cambridge.

(Received June 21, 1932.)

1. When a distortional stress is applied to a solid it produces an instantaneous deformation, which would then remain constant so long as the stress was kept constant if the solid was perfectly elastic. All solids, however, show imperfection of elasticity under sufficiently large stresses, and many do so even when the stress is small. Such imperfection takes the form of a continuous increase of the deformation after the stress has been applied. In some cases the deformation appears to increase indefinitely with the time so long as the same stress is kept on; and in these cases a permanent deformation remains after the stress is removed. In other cases it tends to a finite limit, which is not surpassed unless the stress is further increased; and in many of these, when the stress is removed, the body first returns towards its original state by an amount equal to the original elastic distortion, but then proceeds to creep slowly in the same direction until the whole, or nearly the whole, of the deformation has been recovered. The former type of behaviour, involving permanent deformation, may be called plasticity, the latter elastic afterworking or creep. The distinction is important, but in some accounts of experimental work it does not appear that the stress was maintained long enough to ascertain whether the deformation was tending towards a definite limit, or that the body was sufficiently studied after removal of the stress to find out whether the original state was ultimately recovered; and in consequence phenomena have in some cases been attributed to plasticity that are really due to creep.

The present paper gives first a new derivation of the equations of plastic flow and discusses its relation to some earlier work; second, a theory of creep and a discussion of its relation to experimental evidence; and third, some geophysical applications of the results.

2. The type of plasticity known as elasticoviscosity was expressed in formal equations by J. G. Butcher,* who, however, attributes the underlying ideas to Maxwell. According to this theory the solid consists of two kinds of groups of molecules, one of which is regarded as perfectly elastic, while the other is capable of exerting only a symmetrical pressure or tension. The former type, if the body as a whole is kept under a constant deformation, break down at a definite rate per unit time and are converted into the other type, so that the total shearing stress, measured as an average over all the groups, diminishes spontaneously. On this hypothesis Butcher obtained equations of motion which reduce in extreme cases to those of perfect elasticity and pure viscosity, and these were used by Sir G. H. Darwin in part of his theory of tidal evolution.

These equations may be obtained alternatively as follows. We assume that the rate of increase of strain in an element consists of two parts, one linearly related to the rate of increase of stress, as in an elastic solid, and the other to the stress itself, as in a viscous fluid. If a stress is applied to such a substance, the first part gives instantly an elastic strain, but on account of the second part the strain will increase continually so long as the stress is maintained, and an indefinitely great deformation will ultimately be developed. The phenomenon of plasticity is therefore represented. If u_i is a velocity component, e_{ik} the tensor expressing the rate of increase of strain, and p_{ik} the stress tensor,† we write

$$e_{ik} = e'_{ik} + e''_{ik} \quad (1)$$

to indicate the difference between the elastic and plastic parts of the strains. Now since in an element dp_{ik}/dt is a symmetrical tensor linearly related to e'_{ik} , the rate of increase of elastic strain, we have

$$\frac{dp_{ik}}{dt} = \lambda' e'_{mm} \delta_{ik} + 2\mu' e'_{ik} \quad (2)$$

where λ' and μ' are scalar properties of the material, supposed isotropic. We also introduce the rate of divergence Δ' given by e'_{mm} . Then

$$\frac{dp_{ii}}{dt} = (3\lambda' + 2\mu') \Delta' = 3k' \Delta' \quad (3)$$

Similarly for the rate of plastic strain

$$p_{ik} = \lambda'' e''_{mm} \delta_{ik} + 2\mu'' e''_{ik} \quad (4)$$

$$p_{ii} = (3\lambda'' + 2\mu'') \Delta'' \quad (5)$$

* 'Proc. Lond. Math. Soc.,' vol. 8, pp. 103-135 (1876).

† Jeffreys, "Cartesian Tensors," (1931).

Now as an experimental fact a finite constant symmetrical stress, or hydrostatic pressure, does not produce a change of volume that increases without limit. Hence $\Delta'' = 0$ and $3\lambda'' + 2\mu''$ is infinite. If μ'' also was infinite (4) would evidently make all the components e''_{ik} zero and we should have no plasticity. Hence μ'' is finite. But before using these relations we may use (5) to eliminate Δ'' from (4), giving

$$p_{ik} = \frac{\lambda'' p_{mm}}{3\lambda'' + 2\mu''} \delta_{ik} + 2\mu'' e''_{ik}, \quad (6)$$

$$= \frac{1}{3} p_{mm} \delta_{ik} + 2\mu'' e''_{ik}. \quad (7)$$

Also

$$\Delta = e_{mm} = \Delta' + \Delta'' = \Delta' \quad (8)$$

$$\frac{dp_{mm}}{dt} = 3k' \Delta, \quad (9)$$

$$\frac{\mu'}{\mu''} p_{ik} = \frac{1}{3} \frac{\mu'}{\mu''} p_{mm} \delta_{ik} + 2\mu' e'_{ik}, \quad (10)$$

and adding (2) and (10) we have for a given element

$$\left(\frac{d}{dt} + \frac{\mu'}{\mu''} \right) p_{ik} = \left(\lambda' \Delta + \frac{1}{3} \frac{\mu'}{\mu''} p_{mm} \right) \delta_{ik} + 2\mu' e'_{ik}. \quad (11)$$

Now let $\Delta = d\theta/dt$, so that θ represents the volume expansion of the element; then (9) is equivalent to

$$p_{mm} = 3k' \theta, \quad (12)$$

so that k' is the bulk modulus as usually understood, and the accent may be suppressed. From (11) we now derive

$$\left(\frac{d}{dt} + \frac{\mu'}{\mu''} \right) (p_{ik} - k\theta \delta_{ik}) = -\frac{2}{3} \mu' \frac{d\theta}{dt} \delta_{ik} + 2\mu' e'_{ik}. \quad (13)$$

Then (12) is an equation giving the mean stress, while (13) measures the departure of the stress from symmetry. The mean stress has the same form as in an elastic solid. The departure from perfect elasticity appears in (13) only through the term in μ'/μ'' . These equations are equivalent to those found by Butcher. We can now replace μ' by μ and call it the rigidity, while μ''/μ' , which has the dimensions of a time, is denoted by τ . If τ is large every term in (13) is a derivative with regard to the time, and the equation has the same form as for a perfectly elastic solid. If τ is small in comparison with the time scale of the motion, (13) takes the form appropriate to a fluid of viscosity μ'' or $\mu\tau$.

If the time that elapses is short enough for the total plastic deformation to be small, we can express the deformation E_{ik} in the usual way as

$$E_{ik} = \frac{1}{2} \left(\frac{\partial u_k}{\partial x_i} + \frac{\partial u_i}{\partial x_k} \right), \quad (14)$$

where u_i is the actual displacement, no longer the velocity; and $e_{ik} = dE_{ik}/dt$. Then (13) is equivalent to

$$(1 + Q/\tau)(p_{ik} - k\theta \delta_{ik}) = 2\mu(E_{ik} - \frac{1}{3}\theta \delta_{ik}), \quad (15)$$

where Q is the operator expressing definite integration with regard to the time, following the path of the element.

2.1. It is easily seen that these results can be made to include the theory of plastic flow given by Mises* and Reuss.† It is not necessary to the foregoing argument that μ , or τ , should be constant; it is enough that it should be a scalar. Now such a scalar might be a constant; on the other hand, it may be a function of the state of stress. One scalar derived from the stress tensor p_{ik} is its contracted form p_{mm} ; another is derived from the departure from symmetry, namely,

$$\begin{aligned} & (p_{ik} - \frac{1}{3}p_{mm} \delta_{ik})(p_{ik} - \frac{1}{3}p_{mm} \delta_{ik}) \\ &= \frac{1}{3}\{(p_{22} - p_{33})^2 + (p_{33} - p_{11})^2 + (p_{11} - p_{22})^2\} + 2(p_{23}^2 + p_{31}^2 + p_{12}^2), \end{aligned} \quad (16)$$

and if we transform the axes to the directions of principal stress this evidently becomes $\frac{1}{3}F$, where

$$F = (p_1 - p_3)^2 + (p_1 - p_2)^2 + (p_2 - p_3)^2. \quad (17)$$

Thus τ may be a function of F and p_{mm} . So long as the stress does not depart too much from symmetry most real solids behave as if perfectly elastic, and τ is infinite. When, however, the shearing stresses become great enough, permanent deformation begins, and τ is finite. Mises suggests that this occurs when F exceeds a certain critical value S^2 . This has considerable experimental support for metals, especially in view of the recent work of Taylor and Quinney.‡ When F is greater than S , only a slight excess being possible in practical cases, Mises separates the rate of strain into two parts, as has been done here, and supposes that the part expressing permanent deformation follows the viscous law. The resulting stress equations are equivalent to (12) and (13). Hence if

* 'Z. angew. Math. Mech.,' vol. 8, pp. 161-185 (1928).

† *Ibid.*, vol. 12, pp. 15-24 (1932).

‡ 'Phil. Trans.,' A, vol. 230, pp. 323-362 (1931).

we allow for the possibility that τ may vary with F we can extend the elastico-viscous equations to cover the behaviour of materials with a finite strength. If τ is constant and not infinite, on the other hand, plastic flow begins with any asymmetrical stress, however small, and the equations can represent only the behaviour of materials with zero strength, such as pitch and plastic sulphur.

It appears from (15) that if we know the solution of any problem in elasticity, we can derive the solution of the corresponding elasticoviscous problem by replacing μ by $\mu/(1 + Q/\tau)$, leaving k unaltered.

3. The form (15) shows that if τ is finite and the stress is kept constant, an indefinitely great deformation may ultimately be produced. But in solids showing elastic afterworking or creep the strain in this case tends to a finite limit. If the stress is then removed, there is a partial recovery at once by the amount of the original elastic strain; but then the body gradually changes its shape and finally returns to its original form. This property seems to be absent from single crystals, which behave as perfectly elastic until plastic yield is produced by a sufficient stress; but it is general in crystal aggregates and glasses. It is generally attributed to molecules near the interfaces between crystals; effectively each interface consists of a layer of molecules in equilibrium, each with several possible stable positions, and under non-hydrostatic stress some of these are liable to be pushed over into new positions of equilibrium that are stable in presence of the stress. Thus there is a progressive deformation, until all the molecules that can take up new positions of equilibrium under the stress have done so. This marks the limit of the deformation. If the stress is removed the old positions are again the more stable, and the thermal agitations make the molecules swing back, until finally the whole of the strain has disappeared. In a glass the phenomena are essentially similar except that the mobile molecules are irregularly distributed and not confined to definite surfaces.

3.1. Now for purposes of illustration we may compare this picture with that of a perfectly elastic solid containing limited regions of plastic material. Consider first a long circular cylinder of radius b and rigidity μ , with a core of radius a and rigidity μ' . This is twisted by a given couple applied to the ends. Inertia is neglected. We assume the displacements to be

$$u = -yz\phi, \quad v = xz\phi, \quad w = 0, \quad (1)$$

where ϕ is a function of $r = (x^2 + y^2)^{\frac{1}{2}}$. We find that $\Delta = 0$, and ϕ satisfies

$$\frac{d^2\phi}{dr^2} + \frac{3}{r} \frac{d\phi}{dr} = 0. \quad (2)$$

Then in either medium

$$\phi = A + Br^{-2}, \quad (3)$$

where A and B are constants. The condition that $r = b$ is a free surface gives

$$\frac{d\phi}{dr} = 0, \quad (4)$$

and therefore $B = 0$ in the outer medium. Also $\mu d\phi/dr$ is continuous at $r = a$; hence B is also zero in the inner medium, and ϕ is a constant. Thus any transverse section of the cylinder rotates like a rigid body. The condition for continuity of displacement at $r = a$ is already satisfied.

Then ϕ represents the twist per unit length. The couple applied at the ends is easily found to be

$$N = \frac{1}{2}\pi\phi[\mu(b^4 - a^4) + \mu'a^4]. \quad (5)$$

This solution implies zero stress across the interface between the materials, and suggests that the inner and outer cylinders twist independently. But actually the solution could be correct at the ends of the cylinder only for a special distribution of stress over the ends. For any other distribution there is interaction between the cylinders near the ends, for a distance comparable with b , by Saint-Venant's principle.

Now let the inner material be elasticoviscous. N is taken constant, while μ' is replaced by $\mu'/(1 + Q/\tau)$; or replacing Q by p^{-1} in Heaviside's way*

$$\frac{\mu'}{1 + Q/\tau} = \frac{p\mu'}{p + 1/\tau} \quad (6)$$

$$\phi = \frac{2N}{\pi} \frac{p + 1/\tau}{\mu(b^4 - a^4)(p + 1/\tau) + \mu'a^4p} \quad (7)$$

$$= \phi_0 \frac{p + 1/\tau}{p + 1/\tau'}, \quad (8)$$

where

$$\frac{1}{\tau'} = \frac{\mu(b^4 - a^4)}{\mu(b^4 - a^4) + \mu'a^4} \frac{1}{\tau} \quad (9)$$

$$\phi_0 = \frac{2N}{\pi} \frac{1}{\mu(b^4 - a^4) + \mu'a^4}. \quad (10)$$

The interpretation of (8) is

$$\phi = \phi_0 \left\{ \frac{\tau'}{\tau} + \left(1 - \frac{\tau'}{\tau} \right) e^{-t/\tau'} \right\}. \quad (11)$$

* Jeffreys, "Operational Methods in Mathematical Physics," chap. 1.

When the couple N is applied, the cylinder is at first twisted by ϕ_0 per unit length; and ϕ_0 is simply the elastic twist when the core is taken as perfectly elastic with rigidity μ' . But the twist then proceeds to increase with the time, and ultimately settles down to a steady value $\phi_0\tau'/\tau$, which is the twist when the central rigidity is taken as zero.

If the couple is removed at time t_0 , the subsequent twist is obtained by adding that due to a couple $-N$ applied at time t_0 . The result is

$$\phi = \phi_0 \left(1 - \frac{\tau'}{\tau} \right) e^{-(t-t_0)/\tau'} (1 - e^{-t_0/\tau'}). \quad (12)$$

When the couple is removed there is an immediate reduction of strain equal to the original elastic one; and the remaining strain proceeds to diminish asymptotically to zero. The results, therefore, are in complete qualitative agreement with the observed phenomena of creep. Evidently if a problem has been solved for the case of perfect elasticity, we can allow for elastic after-working by replacing μ by

$$\mu \frac{p + 1/\tau'}{p + 1/\tau}. \quad (13)$$

Then when a stress is applied the immediate distortion is $1/\mu$ times the stress; but the distortion increases to $\tau'/\mu\tau$ times the stress. Then μ is to be interpreted as the rigidity found from the initial displacement, and $\mu\tau/\tau'$, which is smaller, as the rigidity found in a statical test. Two distinct parameters are involved in the law; τ' giving the time-scale of the creep, and the ratio τ'/τ its amount.

3.2. We may notice that the results are greatly affected by the distribution of the elasticoviscous material. Suppose that instead of a cylinder with a central core, we consider a series of cylinders of radius b , fastened together end to end, but some of them with rigidity μ and the others with rigidity μ' . Then the twist in the former is $2N/\pi\mu b^4$, and in the latter $2N/\pi\mu' b^4$. The ends, therefore, turn through an angle

$$\theta = \frac{2N}{\pi b^4} \left(\frac{1}{\mu} \Sigma l + \frac{1}{\mu'} \Sigma l' \right), \quad (1)$$

where l and l' are the lengths of typical cylinders of the two materials. If we write

$$\Sigma l' = m (\Sigma l + \Sigma l') \quad (2)$$

the mean twist is

$$\phi = \frac{\theta}{\Sigma l + \Sigma l'} = \frac{2N}{\pi b^4} \left(\frac{1-m}{\mu} + \frac{m}{\mu'} \right). \quad (3)$$

Now let the rigidity μ' be adapted as before to represent elasticoviscosity. Then

$$\begin{aligned}\phi &= \frac{2N}{\pi b^4} \left(\frac{1-m}{\mu} + \frac{m(p+1/\tau)}{\mu'p} \right) \\ &= \frac{2N}{\pi b^4} \left(\frac{1-m}{\mu} + \frac{m}{\mu'} + \frac{m}{\mu'p\tau} \right) \\ &= \frac{2N}{\pi b^4} \left(\frac{1-m}{\mu} + \frac{m}{\mu'} + \frac{m}{\mu'\tau} t \right).\end{aligned}\quad (4)$$

Thus ϕ starts at the value corresponding to elastic yield, but then proceeds to increase at a constant rate. The behaviour of the composite body in this case exactly resembles that of an elasticoviscous one, with τ replaced by

$$\tau' = \left(\frac{1-m}{m} \frac{\mu'}{\mu} + 1 \right) \tau. \quad (5)$$

Thus the rate of yield, measured by the time needed to double the initial displacement, is less than for a cylinder made entirely of the elasticoviscous material; but its behaviour is still represented by the elasticoviscous law, only with different constants.

The distinction between elasticoviscosity and elastic afterworking may therefore be purely a matter of the distribution of the regions capable of undergoing molecular rearrangement under stress. If the distribution is such that indefinite yield in them would also imply indefinitely large strains in the permanently stable regions, then the yield is resisted by the elasticity of the latter, and provided the strength is not exceeded the composite body will show only creep, which will ultimately disappear after the stress is removed. But if the distribution is such that the weak regions can be greatly deformed without comparable deformations arising in the strong ones, a deforming stress will lead to permanent deformation following the elasticoviscous law. We may say that when the weak regions are enclosed as pockets within the strong ones, the body will show elastic afterworking; when they extend right across the body, it will show elasticoviscosity.

4. Comparison of the law with laboratory experiment shows a qualitative agreement, but there are difficulties in quantitative comparison, because experiments show that the relation is not exactly linear; a linear law can apparently hold only at small strains. Thus Boltzmann* found a creep in glass threads, which were held twisted through 360° for intervals up to 2 minutes. When they were released they did not immediately spring back to

* 'Pogg. Ann.,' Erg.-Bd., vol. 7, pp. 624-654 (1876).

the initial state ; there was a residual displacement, which gradually diminished with time, and practically disappeared in about an hour. The residual displacement was nearly proportional to the time of the original twist, but the recovery did not follow the exponential law. The displacement originally given was in all cases the same, so that there is no means of testing the linearity of the law of creep.

L. N. G. Filon and H. T. Jessop,* working on celluloid, found that the strain under constant stress fitted the formula

$$S = S_0 + at^{1/3} + bt,$$

agreeing with earlier results of Andrade for metals. This evidently represents a form of plasticity. The results fitted no exponential law. In glass the optical double refraction showed no creep or residual effect, though a residual elastic effect had previously been noticed by Jessop alone.† The initial strain is, in all cases, proportional to the stress. The preliminary theoretical discussion given by Filon and Jessop (especially equation (15) on p. 97) leads to the firmo-viscous law for distortion and not to elasticoviscosity or creep as here understood.

In two recent papers on the elasticity of rocks,‡ D. W. Phillips has found that the immediate displacement is accurately proportional to the stress ; but a continued stress produces creep, which is almost completely recovered in time after the stress is removed. The rate of creep is not strictly proportional to the stress ; it increases more rapidly than the stress, reaches a maximum, and at still higher stresses diminishes again. The last peculiarity seems to express the condition where the load does not produce immediate fracture, but does lead to fracture if it is left on long enough. It might be said that yield in the plastic regions throws additional stress on to the strong ones until their strength is exceeded. The maximum creep obtainable was about 0.6 of the initial displacement.

Kelvin§ found that the damping of vibrations in metallic wires was greater the greater the amplitude, and also greater the more vibration the wire has undergone previously. He also found that the rate of damping increased with the speed of the vibration, though the effect was *much less* (Kelvin's italics) than would be expected for friction proportional to the velocity. The

* 'Phil. Trans., A, vol. 223, pp. 89-125 (1922).

† 'Phil. Mag.,' vol. 42, pp. 551-568 (1921).

‡ 'Trans. Inst. Mining Engineers,' vol. 80, pp. 212-242 (1931) ; vol. 82, pp. 432-450 (1932).

§ 'Proc. Roy. Soc.,' A, vol. 14, pp. 289-297 (1865).

vibrations were torsional, and their speed was adjusted by varying the moment of inertia of a hanging body. In these circumstances the equation of motion takes the form

$$A\ddot{\theta} + \mu\dot{\theta} = 0, \quad (1)$$

where A is proportional to the moment of inertia. Then if the time factor is $e^{i\gamma t}$, we have

$$\gamma^2 = \mu/A. \quad (2)$$

Now adapt (2) to allow for creep, by our rule that μ is to be replaced by

$$\mu \frac{p + 1/\tau'}{p + 1/\tau} = \mu \frac{i\gamma + 1/\tau'}{i\gamma + 1/\tau}. \quad (3)$$

Then

$$\gamma = \sqrt{\frac{\mu}{A} \left(\frac{i\gamma + 1/\tau'}{i\gamma + 1/\tau} \right)^{\frac{1}{2}}} = \sqrt{\frac{\mu}{A}} \left\{ 1 - \frac{1}{2} \left(\frac{1}{\tau} - \frac{1}{\tau'} \right) \frac{1}{i\gamma} \right\} \quad (4)$$

for quick vibrations; and the damping factor is $\exp \left\{ -\frac{1}{2}t \left(\frac{1}{\tau} - \frac{1}{\tau'} \right) \right\}$ since to a first approximation $A\gamma^2 = \mu$ is the same for all vibrations. Thus the damping factor should be independent of the speed of the vibration provided this is fast.

But for resistance proportional to the velocity we should take

$$A\ddot{\theta} + 2k\dot{\theta} + \mu\theta = 0, \quad (5)$$

where k and μ are independent of the speed of the vibration. Thus the damping factor is $\exp(-2kt/A)$ or $\exp(-2k\gamma^2 t/\mu)$. The damping is therefore much more severe for the quicker vibrations. Kelvin's observation, so far as it goes, therefore supports the present theory as a first approximation. Unfortunately he does not publish the actual observations for this part of the work, and a closer check is therefore impossible. It appeared that the time needed to halve the amplitude was of the order of some minutes.

The qualitative agreement we have found between theory and observation is enough to indicate that the theory is probably right for sufficiently small strains, but we need an explanation of its inaccuracy for large ones. Now the theory of 3.1 supposed that the elasticoviscous core has a constant radius and a constant value of τ . In an actual solid we should expect that the presence of a distorting stress would increase the tendency of molecules near the flaws to abandon the neighbourhood of their original positions of stability; hence a would be increased and τ reduced. It appears from 3.1 (9) that on both grounds τ' would be reduced. The effect on a is probably of less importance

than that on τ , on account of the smallness of a in comparison with b . The occurrence of a to the fourth power in the equation is a peculiarity of the model actually considered, and when the flaws are generally distributed through the body a lower power of their size may be expected to enter; thus 3.1 (9) probably underestimates the importance of variation in a , though it will probably remain true that it is less important than the associated variation in τ . We should, in any case, expect that the ratio of the rate of creep to the stress will increase with the stress, which is what is required to bring the theory into agreement with experiment.

This can be made somewhat more definite by referring again to Boltzmann's results on the recovery of glass fibres after torsion. The residual displacement, at times more than 2 minutes after the release, was nearly proportional to $1/t$, the time being measured from the moment of release. This empirical result obviously breaks down for small values of the time, and it remains possible that in the first two minutes the exponential law was followed. But let us denote the torsion by θ ; then when there is no applied couple

$$\frac{d\theta}{dt} + \frac{\theta}{\tau'} = 0,$$

and if $\theta \propto 1/t$, $\tau' = t$. Thus τ' may actually vary as the reciprocal of the strain, and the rate of creep for a given stress will contain the strain as a factor.

5. In previous publications I have used an equation of imperfect elasticity suggested to me originally by Sir Joseph Larmor, namely,

$$P = \mu \left(E + \tau_2 \frac{dE}{dt} \right), \quad (1)$$

where E is a component of the deformation and P the corresponding component of distortional stress. This type of behaviour has been called firmoviscosity. If the body is originally unstressed and the stress P is kept constant after time 0, we have

$$E = \frac{P}{\mu} \frac{1}{1 + \tau_2 p} = \frac{P}{\mu} (1 - e^{-t/\tau_2}). \quad (2)$$

This makes the strain tend to a finite limit when t is large, but the immediate strain is zero; E vanishes at $t = 0$. Thus the initial elastic strain is zero. But if we add a constant term P/μ_1 we have again the form 3.1 (11).

With this modification, which removes the outstanding defect of the firmoviscous equation, the law becomes equivalent to 3.1 (8). But the μ in the

firmoviscous law is not the rigidity as determined by the stress needed to produce a given strain either immediately or ultimately; it would be determined by the stress needed to produce a given amount of creep.

6. This result makes it desirable to revise the theory of the transmission of distortional waves that has already been constructed on the basis of the firmoviscous law.* For perfect elasticity the equation of transmission of plane distortional waves is

$$\frac{\partial^2 v}{\partial t^2} = \frac{\mu}{\rho} \frac{\partial^2 v}{\partial x^2}, \quad (1)$$

and if we adapt this to cover elastic afterworking we must write it as

$$\frac{\mu}{\rho} \frac{p + 1/\tau'}{p + 1/\tau} \frac{\partial^2 v}{\partial x^2} = p^2 v. \quad (2)$$

Thus for transmission of a pulse we can write

$$v = e^{-qx} H(t), \quad (3)$$

where

$$q^2 = \frac{p^2}{\beta^2} \frac{p + 1/\tau}{p + 1/\tau'}, \quad (4)$$

β being the ordinary velocity of the waves. Also $\tau < \tau'$.

To interpret, we expand in descending powers of p .

$$q = \frac{1}{\beta} \left[p + \frac{1}{2} \left(\frac{1}{\tau} - \frac{1}{\tau'} \right) - \frac{1}{8} \left(\frac{1}{\tau} - \frac{1}{\tau'} \right) \left(\frac{1}{\tau} + \frac{3}{\tau'} \right) \frac{1}{p} + \dots \right]$$

$$v = \exp \left[-\frac{x}{2\beta} \left(\frac{1}{\tau} + \frac{1}{\tau'} \right) \right] \left[1 + \frac{x}{8\beta} \left(\frac{1}{\tau} - \frac{1}{\tau'} \right) \left(\frac{1}{\tau} + \frac{3}{\tau'} \right) \left(t - \frac{x}{\beta} \right) + \dots \right], \quad (5)$$

for $t > x/\beta$.

There is rigorously no motion till time x/β . Then the pulse enters with a diminished amplitude expressed by the first factor. The second factor proceeds to increase steadily with the time, but evidently is not doubled till a further time of order $\tau'^2 \beta/x$. In this respect it resembles the effect of elastico-viscosity rather than firmoviscosity.

The effect after a long time may be estimated by using the complex integral interpretation

$$v = \frac{1}{2\pi i} \int_L e^{\lambda t - qx} \frac{d\lambda}{\lambda}, \quad (6)$$

* Jeffreys, 'Mon. Not. R. Astr. Soc., Geophys. Suppl.,' vol. 2, pp. 318-323 (1931).

where q is to be found by replacing p in (4) by λ . The saddle points are given by

$$t - \frac{x}{\beta} \left(\frac{\lambda + 1/\tau'}{\lambda + 1/\tau} \right)^{\frac{1}{2}} \left(1 - \frac{1}{2(\tau\lambda + 1)} + \frac{1}{2(\tau'\lambda + 1)} \right) = 0. \quad (7)$$

When rationalized this gives a quartic equation for λ with a real root between $-1/\tau'$ and 0, which is the saddle point. The path of steepest descent bends towards the asymptotic values

$$\lambda = -\infty \pm \frac{ix}{\beta(t - x/\beta)} \left(\frac{1}{\tau} - \frac{1}{\tau'} \right), \quad (8)$$

and the pole at $\lambda = 0$ is to the right of the path and makes unit contribution to the integral. When $\beta t/x$ is large, the saddle-point approaches $-1/\tau'$ and $\lambda t - qx$ approaches $-\infty$. Hence the contribution from the line of steepest descent becomes exponentially small for t large, and in the limit

$$v \rightarrow 1. \quad (9)$$

Elastic afterworking, therefore, cannot convert a sudden impulse into a train of waves. Further, the observed damping, if it was to be attributed to this cause, suggests that

$$\frac{1}{\tau} - \frac{1}{\tau'} = \frac{1}{400 \text{ secs.}}, \quad (10)$$

so that the variation of displacement after the arrival of the pulse is very slow and is not available as an explanation of the blunting of the pulses as they travel.

On the other hand, it has been seen that the effect of scattering, due to minor irregularities of structure, would resemble that of firmoviscosity. We may, therefore, reasonably infer that most of the apparent damping of seismic waves at great distances, and the blunting of the beginnings of the phases, are due to scattering and not to any form of imperfection of elasticity. If elastic after working is present, (10) gives an upper limit to its amount.

7. Consider next the effect on a periodic disturbance, whether free or forced, controlled mainly by rigidity. If the time factor is taken as $e^{i\gamma t}$, we must replace μ by $\mu \left(\frac{i\gamma + 1/\tau'}{i\gamma + 1/\tau} \right)$. This is unity for very quick vibrations, which are therefore unaffected. For very slow vibrations it is again real, though less than μ ; hence these also show no effect of damping, but the effective rigidity for them is reduced. For intermediate speeds it is complex; then forced vibrations will show a phase shift, and free ones will show damping. Thus

we have found a type of imperfection of elasticity that will be in evidence for vibrations of intermediate speeds, but not for very slow or very quick ones. This is what I was seeking unsuccessfully some time ago in trying to reconcile the theory of bodily tidal friction (in the semidiurnal tide) with the observed good transmission of seismic waves and the persistence of the 14-monthly variation of latitude. The success in the meantime of the theory of tidal friction in shallow seas has removed much of the geophysical interest of the problem; but hardly all, since there are bodies such as the moon and Mercury that appear to have been formerly subject to tidal friction and have probably never borne water at the surface. The intrinsic difficulty of handling elastic afterworking quantitatively is the presence of two different parameters. But we have direct evidence that the central core is truly liquid; and the theory of the bodily tide and the 14-monthly variation of latitude, based on the determinations of the rigidity from seismic data and allowing for the fluidity of the core, gives results for these phenomena that are in good agreement with observation.* From this we may infer that either τ' is long compared with both 12 hours and 14 months, or that τ'/τ is near unity, at most about 1.1. In the former case the assumption of perfect elasticity will be valid for tidal phenomena; the same would hold if τ'/τ was exactly 1. But if we assume that creep is important in the theory of the bodily tide we must suppose τ' very short in comparison with 14 months and probably under 12 hours.

Now the phase lag will be α , given by

$$2\alpha = \tan^{-1} \gamma \tau' - \tan^{-1} \gamma \tau.$$

The lag in the bodily tide is under 1° , from the condition that it can account for only part of the secular acceleration of the moon. Hence, either

$$\gamma (\tau' - \tau) < \frac{1}{360},$$

or

$$\frac{1}{\gamma \tau} - \frac{1}{\gamma \tau'} < \frac{1}{360},$$

when $2\pi/\gamma = 12$ hours. If we take the former alternative and put $\tau' = 1.1\tau$ we get $\tau' < 40$ minutes. This is consistent with 4 (10) and leaves open the possibility that appreciable fractions of the observed damping of seismic waves and of the lunar secular acceleration are due to elastic afterworking; though we may be sure that it is not the whole explanation of either phenomenon.

The second alternative makes $1/\tau - 1/\tau'$ so small that no appreciable part of the damping of seismic waves can be due to elastic afterworking. On

* Rosenhead, 'Mon. Not. R. Ast. Soc., Geophys., Suppl.,' vol. 2, pp. 171-196 (1929).

the other hand, if we take the extreme case and make the inequality an equality, the phase lag in the 14-monthly variation of latitude becomes great and we have an inconsistency. On this hypothesis we cannot, therefore, attribute any important damping effect to elastic afterworking.

These geophysical applications have a general interest, because they concern behaviour under very small strains. In an earthquake wave, as usually observed, the displacement is under 0.01 cm. on a wave-length of order 40 km., corresponding to a strain of order 10^{-8} . In the bodily tide the displacement is of order 20 cm. in the earth's radius, again corresponding to a strain of order 3×10^{-8} . These strains are smaller than are observable in most laboratory tests, and for such small values we must suppose that imperfection of elasticity, if it exists at all, follows a linear law.

The Explosive Oxidation of Carbon Monoxide at Lower Pressures.

By G. HADMAN, H. W. THOMPSON, and C. N. HINSHELWOOD, F.R.S.

(Received July 7, 1932.)

The oxidation of carbon monoxide may occur either directly or by an indirect reaction with steam. Under appropriate conditions the oxidation takes place at a measurable rate, and there are also two distinct types of explosive reaction.

The indirect oxidation in presence of steam takes place with measurable speed in the range 550° – 600° C., and appears to involve reaction chains analogous to those occurring in the combination of hydrogen and oxygen.† The direct oxidation requires a considerably higher temperature: in a silica vessel, for example, there is a slow surface reaction at about 700° C. This has been shown to be independent of residual traces of moisture, the kinetics being fundamentally different from those of the indirect oxidation. The direct oxidation can also be brought about by sparking dry mixtures of carbon monoxide and oxygen at high pressures. Bone and Weston‡ have shown that the spectrum emitted by the exploding gases is quite different from that given by the flame of moist carbon monoxide; thus the mechanism of the "dry" reaction is really an independent one. Moreover, Garner§ has shown that there is an abrupt change in the nature of the radiation emitted by a carbon

† 'Proc. Roy. Soc.,' A, vol. 137, p. 87 (1932), where other references are given.

‡ See Bone and Townend, "Flame and Combustion in Gases," chap. 26 (1927).

§ Garner and Roffey, 'J. Chem. Soc.,' p. 1123 (1929); Garner and Hall, *ibid.*, p. 2037 (1930); Garner, Hall and Harvey, *ibid.*, p. 641 (1931).

monoxide flame when a small amount of hydrogen is added, this, again, confirms the existence of two separate mechanisms of oxidation.

This paper deals with a third type of direct oxidation occurring in a region of comparatively low pressure, the extent of which is indicated in fig. 1, where mixtures of carbon monoxide and oxygen combine explosively. The

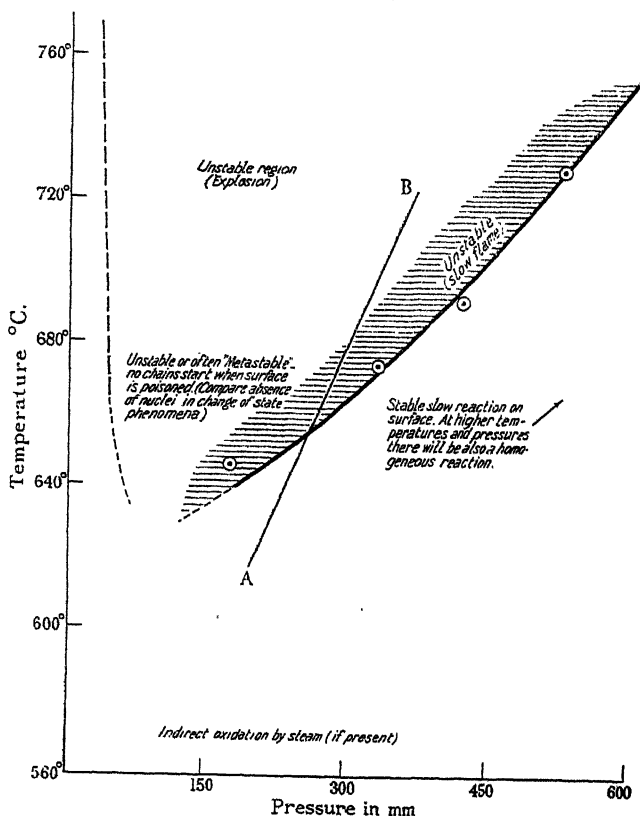


FIG. 1.—Modes of oxidation of Carbon Monoxide at various Temperatures and Pressures.

limits of this region appear to be nearly the same whether the gases have been carefully dried or whether they contain measurable amounts of water vapour. This fact suggests that the mechanism of the explosion depends on the direct interaction of carbon monoxide and oxygen.

With the dry gases the reaction outside the unstable region is very slow indeed, though, if pressure and temperature were raised sufficiently, a region of rapid reaction would presumably be reached eventually, since the velocity increases with both. This region would in turn be bounded by one of thermal explosion. The rapid non-explosive reaction of the dried gases is not easy to

measure, since the limit of the low-pressure explosion region is displaced in the direction of higher pressure as the temperature is raised; thus, experiments in which attempts are made to overcome the slowness of the reaction by using higher temperatures require high pressures, where, for practical reasons, rate measurements are difficult. The high pressure thermal explosion, however, appears in the experiments in which Weston passed powerful sparks through the dry gases and found that they would ignite at a pressure of several atmospheres but not at the ordinary pressure.

The "low-pressure" explosion has many of the characteristics of a reaction depending on the so-called branching chains, but the phenomena are complex and their interpretation difficult. The object of the experiments described in this paper was to study the explosion in more detail and, if possible, draw further conclusions about its mechanism.

Method of Experiment.

Oxygen and carbon monoxide were prepared as described in a previous paper.[†] In the first instance they were dried for 5 weeks over pure phosphorus pentoxide, but in later experiments a much shorter period of drying appeared to be equally effective. All the gases used were contained in spherical globes connected to the reaction vessel by capillary tubing. The taps were lubricated with "Apiezon" grease which has an absolutely negligible vapour pressure. The reaction vessel itself was of silica, and was fixed horizontally in an electric furnace, provided at one end with a quartz window through which any luminescence[‡] of the reacting gases could be observed. The vessel was submitted to prolonged evacuation at bright red heat before use. Pressures were measured by a capillary mercury manometer.

Several methods of observation were adopted, namely:—(1) measurement of the rate of the slow reaction by following change of pressure with time; (2) mixing the gases in a pipette from which they could be displaced by clean dry mercury and gradually compressed in the reaction vessel: this will be referred to as the "compression method"; (3) allowing one of the gases to stream into a known excess of the other already in the reaction vessel and observing whether or not a flash occurs—"mixture method"; (4) preparing in the reaction vessel a mixture at a relatively high pressure and then gradually exhausting, noting the pressure at which ignition may occur—"withdrawal method"; (5) preparing a mixture in the reaction vessel at a low temperature

[†] 'Proc. Roy. Soc.,' A, vol. 137, p. 87 (1932).

[‡] cf. Topley, 'Nature,' vol. 125, p. 560 (1930).

and then gradually raising the temperature, observing the change of pressure and the appearance of any luminescence in the gas (heating method). Observations of the pressure changes and of the flame were supplemented when necessary by analysis of gas samples. It may be said that the interpretation of the results obtained by any one of these methods would have proved difficult without taking into account the information derived from all together.

The Lower Limit.

If the gases are compressed slowly at 650°C . a blue glow is observed to fill the vessel when a certain pressure is reached. The value of this pressure ("lower limit") varies with the condition of the vessel wall, but with a constant technique of working can remain constant through a series of experiments. For example, at 650° in the smallest of the silica bulbs, successive values for an equimolecular mixture were 26, 27 and 27 mm., while after some time further successive values were 32, 32 and 31 mm. Still more drastic variations may occur if the vessel suffers the prolonged action of the gases, so that it is evident that the surface plays an important part in fixing the position of the limit. It should be expressly stated here that there was no systematic dependence upon the time for which the gases had been dried, whether this was 5 weeks or a few days only.

Lower limits of this kind usually depend very much upon the diameter of the reaction vessel, but in the present instance the variations observed with a given vessel were great enough to mask any systematic dependence on the dimensions of different vessels.

The effect of temperature is very small; thus in the larger bulb at 650° values were 13, 13, 12 mm.; at 700° , 15, 14, 13, 13 mm.; then at 650° again, 13, 13 mm.

The influence of variation in the carbon monoxide/oxygen ratio is shown in Table I.

Table I.

Oxygen : Carbon monoxide.	Pressure at limit (mm.).	
	PO_2	PCO
1	16	16
2	19	9.5
3	24	8
6	22	3.7

From this it appears that increase in the amount of oxygen diminishes the amount of carbon monoxide needed to give explosion. This is the opposite of what is found in the thermal explosion of the moist gases, where an increase in the amount of oxygen (leading to smaller reaction velocity) needs to be compensated by an increase in the amount of carbon monoxide if explosion is still to take place.† The contrast confirms the belief that the mechanism of the low pressure explosion is quite different, and may, indeed, be entirely independent of the presence of water.

The most significant fact about the limit is its dependence on the presence of inert gases. Argon and nitrogen lower the partial pressure of the reacting gases required to produce explosion, as shown by fig. 2 and Table II.

Table II.
Larger silica bulb, 650° C. $\text{CO} : \text{O}_2 = 1 : 1$.

Nitrogen : Carbon monoxide.	Pressures at limit (mm.).		
	p_{total}	p_{N_2}	$p_{\text{CO} + \text{O}_2}$
0	19	0	19
2	33	16.5	16.5
4	39	26	13
8	53	43	10
Argon : Carbon monoxide.	Pressures at limit (mm.).		
	p_{total}	p_{A}	$p_{\text{CO} + \text{O}_2}$
0	19	0	19
2	26	13	13
4	30	20	10
8	37	30	7.4

Control experiments showed that the state of the surface did not change seriously during the course of the whole series of experiments.

Other facts must be considered before any attempt is made to discuss the theory of the phenomenon. But it is probably not going beyond legitimate inference from experiment to conclude here that, since the explosion is facilitated by the presence of the inert gas, it depends upon the presence in the gas phase of certain atoms or molecules, which the inert gas hinders from escaping by diffusion.

† 'Proc. Roy. Soc.,' A, vol. 137, p. 87 (1932).

The Upper Limit.

If carbon monoxide is added to oxygen at a pressure greater than about 250 mm. at 650° C. no flash is observed and the gases react with extreme slowness. Under the conditions of our experiments in the smallest silica bulb 300 mm. carbon monoxide and an equal amount of oxygen reacted steadily to the extent of 60 per cent. in 22 hours. When the pressure of oxygen is less than 250 mm. a flash occurs on introduction of the carbon monoxide. Similarly there is a limiting pressure of carbon monoxide above which no flash occurs on the addition of oxygen.

The pressure of oxygen required is rather greater than that of carbon monoxide. An inert gas may to some extent replace either of the reacting gases, *i.e.*, its presence lowers the limit.

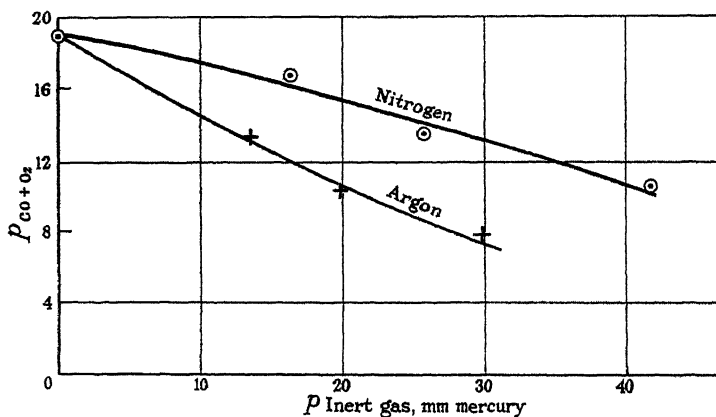


FIG. 2.—Influence of Inert Gases on the Lower Limit.

The limit is independent of the diameter of the vessel, and does not drift in the same way as the lower limit.

Table III.

Temperature 650° C. Addition of carbon monoxide to oxygen.

Diameter of bulb (cm.).	Inert gas present.	Limiting pressure of oxygen for flash (mm.).
1.8	None	240-260
3.2	"	265
4.7	"	240, 250
4.9	"	265
4.9	100 mm. N ₂	205
4.9	100 mm. argon	200

When oxygen was added to carbon monoxide the flashes were fainter and the limit more difficult to determine. In the largest bulb at 650° it was about 170 mm.

The pressure required to prevent the flash from occurring increases rather rapidly with the temperature. If, therefore, a mixture is prepared at a low temperature in the reaction vessel and then heated, *i.e.*, made to pass along the line AB in fig. 1, a flash will occur when the curve representing the boundary of the unstable region is crossed. Since the velocity of change outside the explosion region is very small this "heating method" should give fairly accurate values for the pressure limit. Fig. 1 shows the displacement with temperature of the upper limit so determined with a mixture of composition 1 CO : 2 O₂.

Total pressure (mm.) (measured at the temperature of the flash).	Temperature.
528	730
422	692
333	674
174	646

The explosion limits of moist mixtures have been measured in Semenov's laboratory. The Russian investigators find an upper limit at 650° of 260 mm. for a mixture containing 90 per cent. oxygen, and 190 mm. for one containing 33 per cent. We find 240 mm. for one containing 67 per cent. oxygen. Thus the agreement seems to be rather close. Moreover, Semenov finds the limit displaced from 190 mm. at 650° to about 400 mm. at 680° for the 33 per cent. mixture, while we find for the 67 per cent. mixture a displacement from 240 mm. at 650° to about 360 mm. at 680° . He also finds that increase in the proportion of carbon monoxide lowers the upper limit; this agrees with our observations made by the "mixture method."

The upper limit as indicated by the appearance or non-appearance of the glow seems therefore to be well defined, reproducible, not seriously affected by the degree of dryness of the gases, and independent of the diameter of the vessel. Moreover, it is lowered by the presence of inert gases. It is probably determined, therefore, by an action occurring principally in the gas phase.

Although the limit itself is definite, the degree of completeness of the combustion which accompanies the flash is much more variable and uncertain. When carbon monoxide is added to oxygen at 650° at a pressure of 230 mm. (about 20 mm. inside the limit) a well-defined flash traverses the vessel, but

analysis shows that only a small percentage of the gases has actually been consumed. If the pressure of oxygen is below 200, however, the flash is usually much brighter and endures longer, and analysis shows that most of the gas has been burnt. When oxygen is admitted to an atmosphere of carbon monoxide the combustion accompanying the flash at 650° is nearly always rather small. In general it may be said that the completeness of combustion increases with passage further within the limit, and is determined by factors which seem to vary with the state of the surface of the vessel. There seem to be two sets of factors at work. One set determines the limits between which a flame can be produced in the gas; this flame is an attenuated one and only consumes a considerable proportion of the gas if it can persist. A second set of factors determines the persistence, which must be affected *inter alia* by the power of the vessel walls to reflect reaction chains. The products of reaction probably play an important part in controlling the persistence, and indeed Cosslet and Garner† have found that the presence of carbon dioxide profoundly modifies the percentage combustion of carbon monoxide-oxygen mixtures.

The whole of this question comes into prominence in experiments made to find the limit by the "heating method." When the mixed gases are heated the pressure increases steadily in accordance with Charles' law until the line AB cuts the limit (fig. 1). At this point the normal pressure increase begins to be disturbed by the contraction due to reaction, and, simultaneously, a blue glow appears. As the temperature increases further the pressure diminishes rapidly, the dull flame grows in intensity and finally bursts into a very bright flash; at this moment the manometer registers a sharp kick. With a rate of heating of about 10° a minute the whole phenomenon starts and finishes within a range of about 10° .

At first we interpreted the effect as the setting in of a homogeneous reaction increasing rapidly in rate with temperature and culminating in the ordinary kind of thermal explosion. With no evidence other than that derived from a simple heating experiment it would have been natural to conclude that there was no real crossing of a limit but a continuous transition from slow reaction, through rapid, though still measurable reaction, to explosion. This conclusion, nevertheless, would be misleading, because it would suggest that the non-explosive reaction should be observable over a considerable range of pressure, whereas, in fact, it is only observable in a narrow zone of pressure always immediately adjoining the upper limit. If the gases are heated to the temperature where the glow begins and the manometer shows the setting in of oxidation,

† 'Trans. Faraday Soc.,' vol. 27, p. 176 (1931).

and if at this point either a little more oxygen or a little more carbon monoxide be added the glow is quenched and the reaction rate falls to a negligible value. In our first experiments with the dried gases many fruitless attempts were made to measure what we supposed to be a homogeneous reaction occurring slightly below the explosion temperature; eventually it was realised that the rapid change is inseparably connected with the explosion limit. It is, in fact, a feeble flame, lacking in self-propagating power rather than an isothermal reaction. The explanation of its failure to grow in intensity and consume the gas completely until the temperature is raised some degrees above the point of initiation is evidently the same as that of the incompleteness of the combustion in the "mixing" experiments when the pressure of the gas in the vessel is only just below the limit. The following experiments lend support to this view of the matter. If a mixture of 100 mm. carbon monoxide and 200 mm. oxygen together with 20 mm. carbon dioxide, or 20 mm. nitrogen, is heated, the glow begins at the temperature indicated by the curve in fig. 1, *i.e.*, the small amount of foreign gas has no serious influence on the position of the limit. But on increasing the temperature further the flame does not develop properly and is often quenched again entirely by the carbon dioxide which it itself produces.

It seems to us that in considering the explosion limits the conditions for the first appearance of the flame are the important ones; this is confirmed by the fact that the upper limit so determined is constant and reproducible. The condition that the flame should consume a large proportion of the gas is more difficult to define, and must depend upon factors such as the diffusion to the walls of the vessel of the products which the first chains leave in their wake, and upon the ease of "reflection" of the chains from the walls.

Metastability Phenomena.

Thus far the picture presented is one of two explosion limits with very slow reaction outside them, but this is not a complete picture.

One of the easiest ways of determining the upper limit with a system such as the hydrogen-oxygen system is to mix the gases in the reaction vessel at a high pressure and then slowly to withdraw them, when explosion is observed to occur with great regularity as the limit is passed.

With mixtures of carbon monoxide and oxygen this experiment gives a negative result, whether the withdrawal is rapid or extremely slow. This was shown by repeated trials, the gases sometimes being left for many minutes in the unstable region.

The explanation appeared in the course of a long series of experiments made to determine the influence on the lower limit of varying the proportions of the two gases. Whenever mixtures containing more carbon monoxide than oxygen were used the limit drifted, the explosions became fainter and finally ceased to occur at all. Whenever this happened the *status quo* could be restored by leaving oxygen in the heated vessel for some time.

Thus it appears that exposure of the surface to any considerable partial pressure of carbon monoxide is liable to put it more or less permanently into a state such that no explosion can occur even though conditions of temperature and pressure are favourable. This may be called the metastable state of the system, without prejudging any theoretical issues. Since, in the "withdrawal" experiments, carbon monoxide at a relatively high partial pressure has been introduced, the failure to explode is, obviously, connected with this metastability. It is surprising that the metastability does not affect the "heating" experiments seriously, though it can be understood that a "poisoning" film of adsorbed gas is less likely to survive the disturbing action of increased temperature than that of reduced pressure. There are, indeed, signs that metastability is not entirely absent even here, for, if attempts are made to carry out heating experiments in which the limit is crossed below about 640° , the pale glow may fail to appear at the right temperature, and the limit is overstepped by a variable amount till finally the gas explodes with a very brilliant flash.

Two practical points emerge from these results. If it is desired to map out the explosion area with a minimum of difficulty, the experiments should be carried out at temperatures not much below 650° . Furthermore, to obtain reproducible results in any experiments it is advisable to use mixtures containing excess oxygen, or to heat the vessel with oxygen after each experiment.

Under one special set of circumstances it is possible to obtain the explosion on withdrawal. If oxygen at a pressure just less than the critical value is let into the reaction vessel and carbon monoxide added rapidly, a dull glow begins at once but dies out in a few seconds provided that enough carbon monoxide is added to exceed the limit for the mixture. If, when the glow has become almost imperceptible or, indeed, just imperceptible, the gases are withdrawn, a brilliant flash occurs accompanied by almost complete combustion. The original dull glow leaves behind in the gas phase active products which decay completely if the pressure is kept high, but give rise to explosion if the pressure is rapidly reduced to a value within the limits of the explosion area.

Discussion.

The facts requiring explanation are, first, the existence of a region of explosive combination between two limiting pressures, outside which the rate of reaction is very small, and secondly, the metastable state produced by carbon monoxide on the wall of the vessel.

For brevity it will be convenient here to refer to a discussion given elsewhere† of the various ways in which explosion limits can be explained. The simplest of the theories is that of branching reaction chains. Below the lower limit these are prevented from multiplying effectively because the chain carriers diffuse too rapidly to the wall of the vessel, and are there destroyed. This explains the dependence of the limit on the state of the wall and also the fact that inert gases facilitate explosion, their action being to hinder the diffusion of the chain carriers. The upper limit is governed by some process of recombination of atoms or deactivation of molecules, occurring in the gas phase with greater readiness the higher the pressure. This explains why the upper limit is so much more reproducible, and why it is independent of the vessel and lowered by the presence of foreign gases.

The occurrence of metastability can be interpreted as the occasional complete poisoning of parts of the vessel wall which ordinarily yield by a heterogeneous chemical reaction certain atoms or molecules necessary for the starting of the chains.

Metastability effects are thus connected with the actual initiation of the chain, while the limit phenomena depend on the conditions for the branching of the chains assuming that the mechanism for their starting has operated. There is a rough analogy with the phenomena of change of state, where the presence of nuclei and the appropriate conditions for their growth are both necessary.

As a formal physical theory the branching chain theory seems to account for the known facts, but it may be well to consider briefly some alternatives. In the first place it can be objected that to talk about chains at all is simply a neologistic manner of describing happenings which could just as well be described in purely thermal terms. The lower limit might be regarded as the point at which the local surface oxidation became rapid enough to ignite the gas mixture, and the upper limit as that at which the conduction of heat from the surface through the gas phase became great enough to prevent the ignition. This method of regarding the matter seems to us to be definitely less convenient for two reasons. In the first place, the action of inert gases in facilitating the

† 'Trans. Faraday Soc.,' vol. 28, p. 184 (1932).

ignition in the neighbourhood of the lower limit is difficult to explain by any hypothesis except one depending on the idea that the inert gas prevents diffusion of certain atoms or molecules ; any such hypothesis involves what is essentially a chain mechanism. In the second place, the rate of reaction outside the limits is so very slow that centres on the surface of a very remarkable kind would have to be postulated, capable of giving rise to intense local action without consuming appreciable amounts of the reacting gases. This difficulty could be overcome by the assumption that the centres are only transitorily active while the gases are being mixed and before the true adsorption equilibrium is reached, but even then the influence of inert gases on the lower limit presents a problem, and the fact that the lower limit is independent of temperature would be difficult to explain.

The situation may perhaps be summarised by saying that with the aid of rather special assumptions about catalytically active centres on the surface it is possible to construct a purely thermal theory which leaves only a few of the facts difficult to understand ; while the chain theory gives a simple interpretation of all the facts, which is not only formally correct but, on the whole, more convenient and more suggestive of further experiment. We, therefore, propose to apply it to the oxidation of dry carbon monoxide until the discovery of facts which are inconsistent with it.

It is interesting to note the connection between the metastability phenomena and the theory of Alyea,[†] according to which upper limits are determined by the discontinuous stripping off of an adsorbed film of gas when the pressure falls below a certain value. In this theory the absence of reaction above the limit is attributed to the complete failure of any chains to start, *i.e.*, to the factor which we have supposed to account for metastability in "poisoned" systems. It seems to us impossible that the upper limit can be explained in this way. In the first place, the stripping off of an adsorbed film, as all experience with surface reactions shows, would be a very erratic affair, while the upper limit is not only reproducible but occurs at the same pressure in a silica vessel partly devitrified by long heating to 950° as in one which has suffered no such treatment. Secondly, excess of oxygen or carbon monoxide or of an inert gas will control ignition in the neighbourhood of the upper limit, although it is clearly carbon monoxide alone which can exert "poisoning" influences. Thirdly, the existence of a lower limit is not provided for by the theory.

But the stress Alyea lays on the starting of the chains is important and it seems to us that if the two sets of factors governing the starting of the chains

[†] 'J. Amer. Chem. Soc.,' vol. 53, p. 1324 (1931).

on the one hand, and their propagation on the other hand, are recognised to be independent, most of the facts about this and other reactions can be brought into harmony. The relative importance of the two sets of factors will vary from reaction to reaction.

Passing from the formal physical theory to the detailed consideration of the actual chemical reactions occurring in the chain an element of guesswork inevitably enters. Several different mechanisms have been proposed for the combustion of dry carbon monoxide. A possible cycle of changes will be suggested below, which, although not identical with any so far given by other investigators, incorporates the principal elements of each of them.

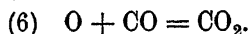
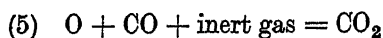
Evidence about the active species concerned in flame reactions can be derived from an analysis of the spectrum. It is not always legitimate to transfer data so obtained to the interpretation of the mechanism of slow reactions, since the latter may not involve the same stages as the flame reactions. In the present example, however, the reaction with which we are concerned is in general an explosive one.

The infra-red radiation from carbon monoxide flames has recently been studied by Garner (*loc. cit.*) and the visible and ultra-violet spectrum by Kondratev.† The conclusion drawn from both series of investigations is that the radiation is emitted by excited molecules of carbon dioxide. The visible spectrum consists of bands, superposed on a continuum, no explanation of the latter having yet been given. Such continua are common in flame spectra, and while their origin is not definitely known, it has been attributed to processes of recombination between free atoms or radicles. The spectroscopic evidence, therefore, indicates the presence in the flame of excited carbon dioxide molecules and possibly of free atoms or radicles which fairly rapidly reunite. Garner has suggested that the processes occurring in the flame involve both "material" chains and "energy" chains.

The experiments described in previous sections indicate that the dry oxidation of carbon monoxide is a surface reaction which readily passes into explosion. If the theory of branching chains is accepted, we must provide a mechanism for the branching. The following cycle of changes seems possible :—

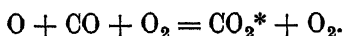
- (1) $\text{CO} + \text{O}_2 = \text{CO}_2 + \text{O}$
- (2) $\text{O} + \text{CO} + \text{O}_2 = \text{CO}_2 + 2\text{O}$
- (3) $\text{O} + \text{CO} + \text{CO} = \text{CO}_2 + \text{CO}$
- (4) $\text{O} + \text{O} + \text{inert gas} = \text{O}_2$

† 'Z. Physik,' vol. 63, p. 322 (1930).

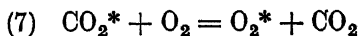


(1) is the primary surface process, (2) provides for the branched chains, (3), (4) and (5) interrupt the chains in the gas, and (6) breaks them at the wall. (6) determines the lower limit. It may also prevent any oxygen atoms leaving the wall at all and so determine the metastable condition of the system. The upper limit is determined by the removal of oxygen atoms in the gas by recombination or otherwise.

The activated carbon dioxide molecules which emit the glow may be formed in processes (3) and (5) and also by an alternative to (2), namely,



Besides the above atomic chain, an "energy chain" such as the following is possible:—



which Garner has suggested (his primary process being the activation of an oxygen molecule).

The existence of oxygen atoms, or some active species which can decay by re-combination, is indicated by the experiment of letting carbon monoxide stream into oxygen so that at the beginning the mixture is just below the upper limit. The pale glow first seen decays in the course of a few seconds. If the mixture is withdrawn quickly, however, a bright flash occurs. Thus it appears that active centres remain and may bring about ignition which would not take place if they were allowed to decay.

On the other hand, the addition of minute traces of iodine was found to inhibit the explosion entirely, and since other experiments give us reason to suppose that the iodine molecule has an exceptional ability to promote transfers of energy, this may indicate the occurrence in the chain of links involving excited molecules also.

We desire to record our indebtedness to the Royal Society, and to Imperial Chemical Industries Ltd., for grants which have made these investigations possible.

Summary.

In the neighbourhood of 700° C. dry carbon monoxide and oxygen combine explosively in silica vessels between certain limits of pressure. Outside these limits the rate of reaction is extremely small. The limits are independent of the degree of dryness of the gases.

The lower limit is affected by the presence of inert gases in such a way as to show that it is governed by the diffusion of active centres of some kind to the vessel wall. The upper limit is independent of the vessel, and affected by foreign gases in such a way as to show that it is governed by gas phase deactivation.

The lower limit is independent of temperature; the upper limit moves rapidly towards higher pressures as the temperature rises.

The action of carbon monoxide on the wall of the reaction vessel may put the system into a state where no explosions are possible; this may be called the metastable state.

The experimental results are interpretable in terms of the theory of reaction chains, which is adopted as the most convenient after a consideration of other possibilities. The explosion limits are determined by the conditions for the propagation of the chains, while metastability depends upon failure of any chains to start. It is pointed out that, just as in the phenomena of change of state, the conditions for the initial production of active centres, and those for their survival and multiplication are independent. The existence of two sets of factors accounts for the complexity of the experimental facts.

The mechanism of the chemical changes involved in the chains is discussed.

The Lower Pressure Limit in the Chain Reaction between Hydrogen and Oxygen.

C. N. HINSHELWOOD, F.R.S., and E. A. MOELWYN-HUGHES, D.Phil.

(Received July 19, 1932.)

Previous investigations* have shown that above 460° C. there are two pressure limits between which the combination of hydrogen and oxygen takes place explosively and outside which it may abruptly become extremely slow. Practically difficulties have hitherto hindered a detailed study of the conditions governing the position of the lower limit. A modification of the earlier experimental arrangement has allowed these difficulties to be overcome, and the influence on the limit of hydrogen-oxygen ratio, vessel diameter, temperature and presence of inert gases has been investigated. The influence of these

* 'Proc. Roy. Soc.,' A, vol. 122, p. 610 (1929); vol. 130, p. 640 (1931); vol. 134, p. 1 (1931).

various factors corresponds in general to that found in other reactions which exhibit limit phenomena. The results are consistent with the hypothesis, already put forward and discussed, that the lower limit is that pressure where reaction chains cease to be broken at the walls of the vessel as rapidly as they are started by the "branching" of existing chains.

The method of experiment was simply to prepare in A (fig. 1) a mixture of appropriate composition, measure off a small proportion of it in the capillary pipette B, and then to allow the contents of B to stream into the evacuated silica reaction vessel C. The coefficient of sharing between B and C was determined by a special calibration. C was heated in a horizontal electric

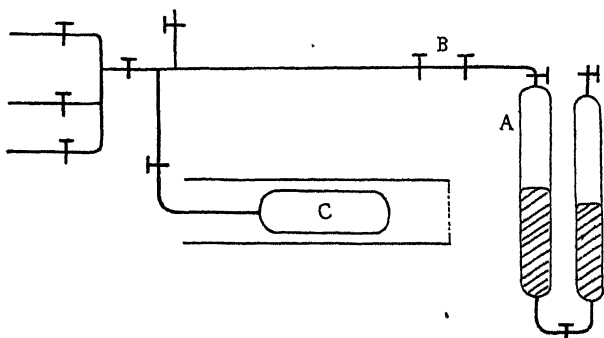


FIG. 1.

furnace provided with a quartz window. If the calculated final pressure in C was below the limiting value nothing was seen when a pipetteful of mixture was added, while if it was above the limit a flash was observed. The temperature of C was measured with a thermocouple.

The position of the limit could be determined fairly sharply, as the following example shows:—Pressure in pipette 95, 100, 105 mm.—no flash; 109, 115 mm. and higher—flash; limit 107 mm., corresponding to a pressure in the reaction vessel of 3.38 mm.

The flashes were feeble and had to be observed with a dark-adapted eye. The position of the limit tended to shift with continued use of the reaction vessel, as might be expected from the fact that the whole phenomenon depends upon surface deactivation. It was found, however, that reasonably constant and concordant results could be obtained from day to day if the reaction vessel was washed out with oxygen for a few minutes before each experiment. In one of the bulbs the following limits were found at intervals during about 3 weeks: 4.20, 4.25, 3.33, 3.52, 4.01, 4.50, 3.20, 3.00, 3.30, 4.20, 3.34, 3.31.

Influence of Temperature.

From 550° to 650° the influence of temperature is negligible. At higher temperatures the furnace is too bright for the luminescence accompanying the explosion to be observed properly.

Thompson and Hinshelwood* found in studying the upper limit that no explosions occurred at all below about 450°. They assumed this to be the temperature at which the upper and lower limits coincide, and inferred that the lower limit was at about 8 mm. This value is considerably higher than the values found for the lower limit between 550° and 650° in the present series of experiments; thus it appears that below 500° the lower limit ceases to be independent of temperature. This conclusion has been confirmed by direct measurement. Fig. 2 shows that the curves giving the variation of the two

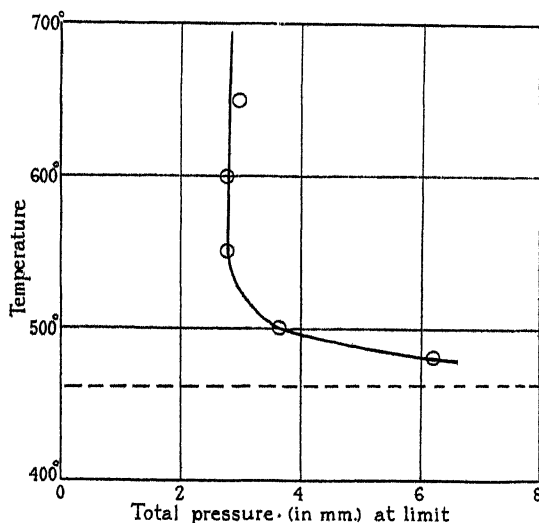


FIG. 2.—Influence of temperature on the lower limit. The dotted line represents the temperature at which the upper and lower limits were previously found to coincide.

limits with temperature do not cut sharply as Hinshelwood and Thompson's approximate calculation virtually assumed, but that the curve for the lower limit bends round at the lower temperatures quite markedly.

Silica Bulb, 1.8 cm. diameter.

Temperature	650°	604°	550°	500°	480°
Lower limit (mm.)	2.99	2.74	2.75	3.62	6.20

* 'Proc. Roy. Soc.,' A, vol. 122, p. 615 (1929).

Influence of the Relative Proportions of Hydrogen and Oxygen.

The results in Table I were obtained in the bulb of 1.8 cm. diameter at 550°.

Table I.

p_{H_2}/p_{O_2}	Critical pressure (total).	p_{H_2} (mm.).	p_{O_2} (mm.).
4:1	3.9	3.12	0.78
2:1	3.45	2.30	1.15
1:1	3.3	1.65	1.65
1:2	3.2	1.07	2.14
1:4	4.3	0.86	3.44

If p_{H_2} is plotted against p_{O_2} the curve obtained is not quite a rectangular hyperbola, but approximates roughly to one.

A repetition of the experiments in a larger vessel where the pressures were smaller gave less reproducible results, and the curve of p_{H_2} against p_{O_2} showed a more marked departure from symmetry, in the sense that the product $p_{H_2} \cdot p_{O_2}$ increased considerably as p_{H_2} decreased.

Influence of Inert Gases.

Inert gases hinder the diffusion of the chain carrying species to the wall and thus facilitate explosion. Accordingly the partial pressures of the reacting gases at the limit are diminished by the presence of the inert gases. In accordance with the difference in the diffusion coefficients argon exerts a much greater influence than helium. All the results recorded in Table II refer to the bulb of 1.8 cm. diameter and to a temperature of 550°. The ratio $H_2 : O_2$ was 2:1.

Table II.

Nitrogen.		Carbon dioxide.	
p_{N_2}	$p_{(H_2+O_2)}$	p_{CO_2}	$p_{(H_2+O_2)}$
0	3.20	0	4.50
2.12	2.12	6.24	3.12
3.10	1.55	11.4	2.79
4.94	1.65		
5.25	1.05		
6.15	0.88		

Table II—(continued).

Helium.		Argon.	
P_{H_2} .	$P(H_2 + O_2)$.	P_A .	$P(H_2 + O_2)$.
0	3.33	0	4.25
5.45	2.73	2.90	2.90
8.70	2.18	4.41	2.21
(0	3.52)	6.72	1.68
		6.51	1.09
		(0	4.20)

From series to series the control experiments without inert gas vary somewhat. In a given series the variation is not serious, as may be seen from the initial and final experiments of the series with argon and with helium.

Influence of Vessel Size.

The limit at 550° for a mixture of composition $2H_2 : O_2$ in a number of silica bulbs is recorded in the following table. As was to be expected, vessels of the same dimensions differ from one another owing to qualitative differences in their surfaces. Nevertheless the greater ease of explosion in the larger vessels is quite evident.

Table III.

Diameter of bulb (cm.).	Pressure at lower limit (mm.).		Observations.
5.4	0.76	Mean 1.06	Length 9 cm.
5.4	1.36		Apparently identical with first.
4.9	0.91		Length 9 cm.
3.2	2.85		"
1.8	3.64		"

The values in Table III are the mean of several determinations made after the surface of the vessels had been conditioned by a little use.

Discussion of Results.

If the chain mechanism involves hydrogen and oxygen in a symmetrical way, and if the chains are broken at the wall of the vessel, then it can

be shown* that for a given vessel and with a given inert gas the expression

$$p_{H_2} \cdot p_{O_2} \left(1 + \frac{p_{\text{inert gas}}}{p_{H_2} + p_{O_2}} \right)$$

should be constant. The theory is based upon a number of simplifications, however, and can only be expected to hold approximately.

As shown above, in the absence of inert gas there is a roughly hyperbolic relationship between p_{H_2} and p_{O_2} as there should be. For the influence of inert gases we should find that $1/(p_{H_2} \cdot p_{O_2})$ plotted against $1 + \frac{p_{\text{inert gas}}}{p_{H_2} + p_{O_2}}$ gives a straight line, assuming that the inert gas only acts by hindering diffusion to the walls. As Melville and Ludlam† point out, the slope of the line will be proportional to the diffusion coefficient of the active chain carriers through the inert gas. Fig. 3 shows the results for helium,

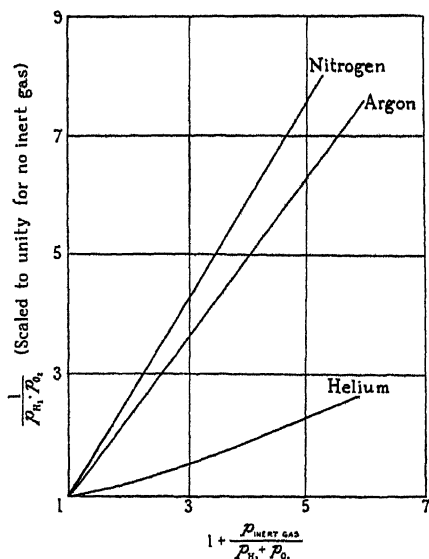


FIG. 3.—Influence of inert gases on the lower limit.

argon and nitrogen. The relative slopes for argon and helium are about what would be expected. Carbon dioxide exerts considerably less effect than it should, presumably because of a specific deactivating influence

* Compare the derivation given for the analogous example of phosphine and oxygen, 'Proc. Roy. Soc.,' A, vol. 125, p. 294 (1929).

† 'Proc. Roy. Soc.,' A, vol. 132, p. 108 (1931).

which it exerts on the chain carriers, or of an influence on the nature of the surface.*

In the ideally simple chain reaction with wall deactivation the product $p_{H_2} \cdot p_{O_2}$ would vary inversely as the square of the diameter of the vessel, assuming the latter to be an infinitely long cylinder. The table in the previous section shows irregular variations, due to the fact that the wall does not remove the chain carriers with perfect efficiency, and that the actual efficiency varies with the exact condition of the surface. It is evident that $p_{H_2} + p_{O_2}$ varies roughly in an inverse linear ratio with the diameter for vessels of the same length; thus the product $1/p_{H_2} \cdot p_{O_2}$ will vary as something of the order of the square of the diameter, but a more detailed analysis is obviously useless.

It may be said that the results generally are consistent with the theory of a lower limit governed by the diffusion of active centres to the wall, but that various specific actions make the agreement with the quantitative theory only approximate.

We wish to express our indebtedness to Imperial Chemical Industries, Ltd., and to the Royal Society for grants with which apparatus for these experiment, was obtained.

Summary.

The influence of temperature, hydrogen-oxygen ratio, presence of inert gases and vessel diameter on the lower limit of the low pressure explosion region in the combination of hydrogen and oxygen has been investigated in more detail than was possible in earlier work. The results are consistent with the hypothesis, previously advanced and discussed that the lower limit is determined by a balance between the branching of reaction chains and the removal of chain carriers by diffusion to the wall of the vessel.

* Carbon dioxide exerts some deactivating influence in the gas as shown by its influence on the upper limit; but the effect is only of the same order as that of other gases. The following experiments to test this may be recorded here:—

540° upper limit (mm.).	P_{total} .	$P(H_2 + O_2)$.	P_{CO_2} .
$H_2 : O_2 = 2 : 1$	73	73	0
	62	47	15
	65	40	25
	59	30	29

The γ -Rays of Thorium B and of the Thorium C Bodies.

By C. D. ELLIS, F.R.S.

(Received August 6, 1932.)

[PLATE 12.]

§ 1. *Introduction.*

A word of explanation is necessary in presenting this paper. It is unfinished, but designedly unfinished. When β -ray spectra were first discovered, the papers describing the experiments made no claim to give complete lists of the lines, publication was justified by the discovery of even a few strong lines. Then came the period when the technique had reached a sufficient stage of development to detect with fair certainty all except the faintest lines, and further, and this is the important point, the attainable accuracy was more than sufficient for all theoretical purposes, since in reality but little theory existed. From this stage we are just emerging. Recent experiments have proved beyond doubt that the γ -rays are associated with transitions between α -particle stationary states in the nucleus, and it is clear that the β -ray spectra will play the same rôle in giving information about these states as have the optical X-ray spectra for the electronic states. Yet this information will not be obtained in its final form by one experiment. Far more we have to expect a gradual accumulation of knowledge in which step by step the exact energies of the lines, their allocation to the correct body, and their intensities are obtained to a sufficient accuracy.

The present paper is a contribution to our knowledge of the thorium ($B + C + C' + C''$) spectra,* and for the reasons which have been mentioned the treatment is neither exhaustive, nor are the results equally detailed for all parts of the spectrum. It is hoped, however, that the results are presented in a form which will prove accessible, and for this reason only slight reference is made to experimental technique. A detailed study, which proved very laborious, has been made of the intensities of the groups. Beyond mere visual estimation, no attack had previously been made on this problem, which I suggest is of equal importance to that of finding the energies of the groups. In the course of this work many new lines were detected, and some discrepancies

* For previous work see "Radiations from Radioactive Substances," Rutherford, Chadwick and Ellis, p. 368 (Camb. Univ. Press).

with previous workers noted. Careful measurements were therefore made of the energies of all the lines. The final results constitute an entirely independent investigation of the energies and intensities of the groups in these spectra.

The arrangement of the paper is as follows. The next section contains Table II, which is a list of the energies and intensities of all the β -ray lines found from a source of thorium ($B + C + C' + C''$) in transient equilibrium, and various points of importance in connection with the deduction of these final figures are there discussed. The succeeding sections deal in turn with the division of these lines among the various bodies, and the deduction of the γ -rays from the β -ray spectra. The assignment of lines to ThBC is based on Black's* work, that of the lines to the other disintegrations follows in the main the evidence obtainable from the α -particle groups. The details will be found

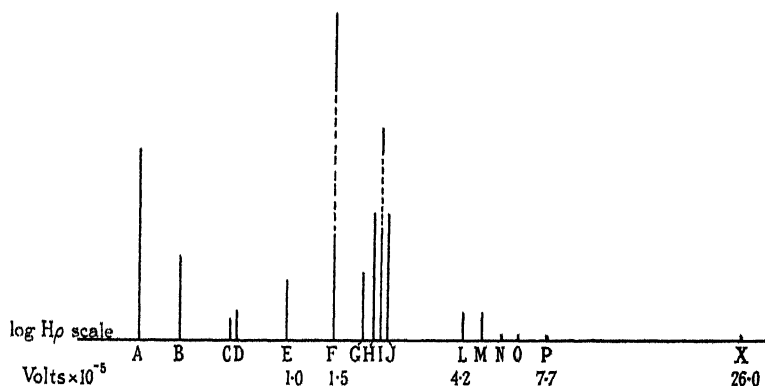


FIG. 1.—Strong lines of Th ($B + C + C' + C''$) β -ray spectrum.

in the sections devoted to the different disintegrations. It must be emphasised that this allocation of the lines is in some cases doubtful, but it is convenient to come to a provisional decision provided always that the results are so presented as to make it easy to alter this when better data are obtained. An important step in this direction is to name the lines by letters so that their identity may persist even if their origin or the value of their energies are altered. The system of nomenclature adopted here was explained in a letter to 'Nature.'† To facilitate identification a map of the stronger lines is shown in fig. 1. The height of the lines in the diagram shows very approximately the relative photographic intensity of the different groups. A scale of $\log H_p$ is used since this puts the lines in the same relative positions as they appear

* 'Proc. Roy. Soc.,' A. vol. 109, p. 166 (1925).

† Ellis, 'Nature,' vol. 129, p. 276 (1932).

on the photographic plate. For suppose the limiting radii of curvature that can be used in the apparatus are ρ_1 and ρ_2 , then in a magnetic field H a section of the spectrum $H\rho_1$ to $H\rho_2$ falls on the photographic plate which is, of course, of the same length in all experiments. But equally in the diagram a region $(H\rho)_1$ to $k(H\rho)_1$, where k is a constant, will occupy the same length $\log k$ whatever the value of $H\rho_1$. With limiting radii of 6.3 and 9.3 cm. it will be found that the portion of the spectrum which could be photographed in any one spectrum corresponds to a length in fig. 1 a little greater than F to J. Portions of these spectra have also been investigated recently by Sze-Shih-Yuan.* His results appear to be in general agreement with the present measurements, but a detailed comparison is difficult at present, since he uses a different absolute scale of energies and it is not clear what value of e/m_0 is employed. He further does not measure the intensities.

Some reproductions of the photographs obtained are shown in Plate 12. The photograph, fig. 5, Plate 12, shows the complexity of the lines in the region Bd-E;† most of these lines are due to X-rays generated in the radioactive atom by conversion of the γ -rays. The photographs figs. 6, 7, 8 and 9 taken with a finer source, an activated platinum wire 0.06 mm. diameter. In fig. 6 the exceptionally strong line F will be noted; its breadth is due to over-exposure. There are satellities to both the lines E, figs. 5 and 6, and I, fig. 7, which owing to the fineness of the lines are clearly separated. Fig. 8, Plate 12, is interesting since it shows the β -ray line X due to the powerful high-frequency γ -ray of $2\frac{1}{2}$ million volts from thorium C'' . Pb.

The actual shape of the lines was found by measuring the density with a recording photometer, and owing to the fineness of the lines it was possible to deduce that the distance between the foot and the peak did not differ from that to be expected from a strictly homogeneous group by more than that corresponding to a change of one part in about three thousand of the total energy. On general grounds one would expect the electronic groups to be homogeneous to something of the order of one part in one hundred thousand, but these measurements suffice to show that there is no gross inhomogeneity. In one particular case, however, this information is of real importance. There are two γ -rays emitted by thorium C'' . Pb which will be called γ_L and γ_M , since by conversion in the K level they give the lines named L and M. These γ -rays have been identified by the excited spectrum method, that is by the

* 'C.R. Acad. Sci. Paris,' vol. 194, p. 874 (1932).

† To avoid confusion only a few lines are lettered, the lines C and D are the right hand ones of the two doublets on the left of the photograph.

photo-electric groups they eject from heavy metals, both by Thibaud* and by the writer in the present series of experiments, so that there is no doubt that both exist in emission. The curious feature, however, is that the quantum energies of these γ -rays differ by exactly the same amount as the K and L_I absorption energies of the electronic system, so that the β -ray line due to the L_I conversion of γL coincides apparently exactly with the K conversion of γM . The β -ray line M is therefore double and this view is supported by the intensities. It is both probable on theoretical grounds and supported by experiment that a β -ray line due to K conversion is about seven times as strong as the corresponding line due to L_I conversion. Now the β -ray line L has an intensity 1.7 in arbitrary units, so that we should expect the L_I conversion to be about 0.25. Exactly at the place where this line should be there is the strong line M of intensity 1.7, which we therefore suppose to be made up of 0.25 by the L_I of γL , and 1.45 by the K of γM . The L_I line of the γM should then also be about 0.2. There is a β -ray line named N which has just the right energy for this purpose and its intensity was measured as 0.3, which, considering the difficulty of the intensity measurements, is sufficient agreement. It is interesting to note that both these γ -rays appear to behave normally as regards the intensity of conversion, as we have in fact assumed, since the β -ray lines due to conversion in the M electronic levels [viz. the lines Ma and Na (see Table II)] both have intensities about 1/24 of the K conversion which is the usual figure found.

Since it is then clear that the β -ray line M is double it is a matter of some interest to see whether it is slightly wider than L. These two lines are shown in fig. 9, Plate 12, it is at once obvious that they look similar and actually the photometer records suggest that they do not differ in shape by an amount corresponding to a change of 1 part in 5000 in the energy. Even taking into account that the two superimposed lines differ greatly in intensity I find it difficult to believe that they differ in energy by as much as 1 part in 2000, and actually the superposition looks to be exact. While there is no other case so well authenticated as this, there are indications that there are other γ -rays showing this curious relation that their energies differ by an amount exactly equal to energy differences of the electronic structure. It may be only a coincidence and certainly further investigation is needed before this observation would justify an attempt at a theoretical interpretation.

The radioactive source used in these experiments was thorium B in transient equilibrium with its products. The scheme of transformation is shown in

* 'Thesis,' Paris (1925).

fig. 2 from which it appears that we are concerned with five different types of disintegration, viz., thorium B.C, thorium C.C'', thorium C.C', thorium C''.Pb and thorium C'.Pb. It has been possible to identify β -ray groups associated with γ -rays from thorium B.C, thorium C.C'' and thorium C''.Pb. Particular interest attaches to the case of thorium C.C'', since this is the disintegration for which a series of α -ray groups have been detected by Rosenblum* and which provides one of the best examples of the association of γ -rays with the nuclear α -particle levels. The measurements of the present paper, and the latest results of Rosenblum and Valadrest† are carefully compared from this point of view. The detailed discussion will be found in section 4 and the conclusion is reached that the new measurements provide convincing evidence in favour of the connection between γ -rays and α -particle levels. Little need

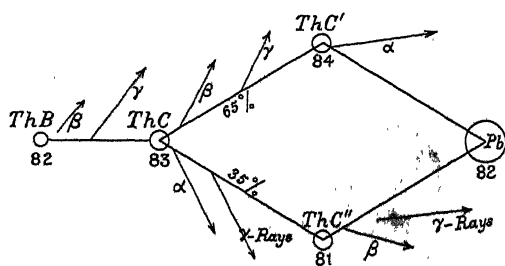
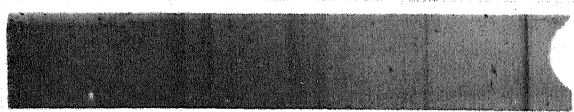


FIG. 2.

be said about the γ -rays of thorium B.C and thorium C''.Pb. Both are β -ray disintegrations and their γ -ray emission has been studied by many investigators. The cases of thorium C.C' and thorium C'.Pb are more interesting. Thorium C'.Pb is an α -ray disintegration, and following Gamow's theory of the γ -rays from the α -ray disintegration of thorium C.C'', we should only expect γ -rays in detectable amounts if the α -particles were emitted in a series of groups whose intensities did not differ by a factor of more than one or two hundred, and further the highest energy group would be expected to be one of the strongest. Nothing like this is observed, so presumably there are no γ -rays. It is well known, however, that thorium C'.Pb emits groups of long range α -particles analogous to the radium C.C' disintegration and this would lead us to expect γ -rays from thorium C.C' corresponding to the energy differences of the α -particle groups. There is some slight evidence in the β -ray spectrum for some of these γ -rays (Section 6), but I have not obtained satisfactory evidence of what would be expected to be the most prominent

* 'J. Phys. Rad.', vol. 1, p. 458 (1930).

† 'C.R. Acad. Sci. Paris,' vol. 194, p. 967 (1932).



Bd Dg E

FIG. 5.



E Eb Ec F

FIG. 6.



G H I J

FIG. 7.



X

FIG. 8.



L M

FIG. 9.

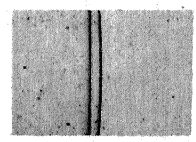
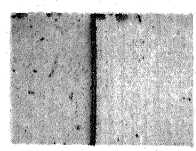
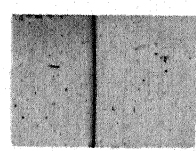


FIG. 10.

transition. Skobel'tzyn,* however, investigating the γ -rays by the Compton effect, appears to have detected this γ -ray, and it would therefore seem that there is here a case of very slight internal conversion. There is a considerable amount of information in these β -ray spectra about internal conversion and it is proposed to consider it separately in another paper. The result that the L_I conversion line is about one-seventh the intensity of the K conversion is supported by these measurements. We have already used this in discussing the superposition of two β -ray lines, and it proves generally to be of great help in disentangling the spectra.

There is one region Bd-Df2 in which there is a series of close lines. A reproduction of a photograph of these lines is shown in fig. 5, Plate 12. It seems certain that most of these lines arise entirely from the electronic structure after the ionisation of the K shell by the conversion of the γ -rays and they can be considered as due to the conversion of the K X-rays in the various electronic shells. They will not be further considered in the present paper since the best method of analysing this region will be to compare it with the corresponding region in the radium (B + C) spectrum.

§ 2. *Experimental Results.*

The usual type of semicircular focussing apparatus was used, the radii of curvature of the extreme rays being between 6.3 and 9.3 cm. In the latter part of the work a large permanent magnet† was used which proved to be exceptionally convenient and contributed greatly to the accuracy of the work.

The work was originally started with a view to finding the intensities of the groups in these spectra, since the energies had already been studied by several workers. It soon appeared, however, that the values of the energies obtained from my plates did not agree sufficiently well with those already published, and it was therefore necessary to make a study also of the energies of the groups. In the greater number of the experiments measurements were made only of the relative velocity of the groups, that is, of the relative values of $H\rho$. Provided the magnetic field is uniform, the relative values of $H\rho$ of two groups can be obtained simply from the relative values of the radii of curvature when they are photographed in the same magnetic field, and thus by a stepwise method of comparison it is possible to compare in turn all the groups. By a proper attention to technique it was found possible to obtain values for this ratio which in the case of the stronger lines were repeatable to 1 part in 3000. These values were obtained with two different sets of apparatus using two

* 'C.R. Acad. Sci. Paris,' vol. 194, p. 1486 (1932).

† Cockcroft, Ellis and Kershaw, 'Proc. Roy. Soc.,' A, vol. 135, p. 628 (1932).

different magnets. Taking into account various other sources of error, it still seems probable that the ratio of the $H\rho$'s of neighbouring lines is correct to 1 part in 1000. An important feature in this part of the measurement was to be sure that the magnetic field was homogeneous over the part used. The polepieces were 28 cm. by 17.5 cm. and only the central part of the field was traversed by the β -particles. The usual method of checking the uniformity of the field, which is to see whether the same ratio is obtained for two lines when they are moved to different parts of the plate, was not considered to be sufficient, and the following procedure was adopted. In the first place the theoretical lack of homogeneity was calculated on the assumption that the polepieces were infinitely permeable, that is to say, that the magnetic field was similar to an electrostatic field. The applicability of this calculation was then tested by the following experiment. The apparatus was placed in position and a photograph taken of certain strong lines. Then, without moving either the source or the photographic plate, the whole apparatus was raised a small distance of 1 to 3 cm. the magnetic field remaining constant throughout. Another photograph was then taken of the same lines with the box in this raised position. The β -particles will now go through the more outlying parts of the magnetic field where the field is weaker and they will therefore traverse slightly longer paths. By using a very fine source, it was possible to observe the lines of the second exposure separated from those of the first exposure. This separation can be measured and compared with that to be anticipated from the calculated inhomogeneity. In fig. 10, Plate 12, will be seen three photographs showing the results obtained when the box was raised respectively 1.5, 2.5 and 3.5 cm. In the first the two lines cannot be seen separated, whereas in the two latter the lines are clearly separated. The results for this and other lines are shown in Table I, from which it will be seen that the measured separation agrees excellently with the calculated separation.

Table I.

Distance the box was raised from standard position in cm.	ρ of line investigated in cm.	$\delta\rho/\rho$ measured.	$\delta\rho/\rho$ calculated.
1.5	7.18	ca. 0.0005	0.0009
	8.27	0.0030	0.0026
	8.78	0.0046	0.0046
	9.09	0.0065	0.0067
2.5	7.18	0.0030	0.0034
3.5	7.18	0.0103	0.0105

These measurements were carried out with β -particles of an average H_p of about 1500, corresponding to 1.7×10^5 volts. While the homogeneity will be different owing to the changing permeability of the material of the pole-pieces, the lack of homogeneity over the region used is so small that the correction obtained from this calculation may safely be applied to all photographs. Even when the extreme radius of curvature is used, the correction is only about 1 part in 1000.

The direct experimental results representing the relative H_p 's of the lines are shown in column III of Table II and are given as the logarithm of the H_p .

Table II.— β -ray Groups from a source of Th B in Transient Equilibrium with its Products.

I. Name.*	II.† Origin.	III.‡ log H_p .	IV. H_p .	V. Energy.	VI.§ Photographic intensity.	VII. True intensity.
A	C. C''	73320	541.0	0.2514	} Not measured.	
A1		73384	541.8	0.2521		
Aa		73910	548.4	0.2581		
Ab		75610	570.3	0.2787		
B		82007	660.8	0.3708		
Ba		82217	664.0	0.3743		
Bb		83721	687.4	0.4001		
Bc		84136	694.0	0.4076		
Bd		91721	826.4	0.5693	0.60	0.60
C	B. C	92006	831.9	0.5765	1.05	1.00
C1		92161	834.9	0.5804	0.20	0.20
C2		92505	841.5	0.5892	0.30	0.30
Ca		92903	849.2	0.5994	0.80	0.70
D		93155	854.2	0.6061	1.45	1.30
Da		94004	871.0	0.6289	0.80	0.70
Db		96290	918.1	0.6946	0.45	0.35
Dc		96547	923.6	0.7023	0.45	0.35
Dd		96803	929.0	0.7101	0.35	0.30
De		97392	941.7	0.7285	0.60	0.45
Df		97584	945.9	0.7346	0.45	0.35
Df1		98328	962.2	0.7587	0.10	0.08
Df2		98600	968.3	0.7674	0.10	0.08
Dg		01160	1027.1	0.8566	0.50	0.40
E		04398	1106.6	0.9830	3.00	2.30
Ea		04549	1110.4	0.9892	0.30	0.20
Eb		07207	1180.5	1.1066	0.80	0.75
Eb1		07884	1199.1	1.1386	0.25	0.20
Eb2		08857	1226.2	1.1856	0.05	0.40
Ec		09729	1251.1	1.2295	0.40	0.30
Ecl		10510	1273.8	1.2699	—	—
Ed		13699	1370.8	1.4485	0.25	0.20
F		14170	1385.8	1.4766	200	165
F1		15492	1428.6	1.5584	0.05	0.04
Fa		16034	1446.6	1.5932	0.40	0.35
Fb		16710	1469.3	1.6373	0.70	0.60
Fb1		16997	1479.0	1.6563	0.07	0.06
G	C''. Pb	20243	1593.8	1.8868	3.60	3.20

Table II—(continued).

I. Name.*	II.† Origin.	III.‡ log H ρ .	V. H ρ .	V. Energy.	VI.§ Photographic intensity.	VII. True intensity.
G1		20732	1611.8	1.9241	—	—
Ga	C. C''	21928	1656.8	2.0173	0.70	0.65
H	B. C	22815	1691.0	2.0892	6.60	6.00
H1	C. C''	23324	1710.9	2.1313	—	—
I	B. C	24329	1751.0	2.2170	25.0	23.3
Ia	B. C	24402	1754.0	2.2235	1.6	1.5
Ia1	B. C	24596	1761.8	2.2403	0.08	0.08
Ia2	B. C	25627	1804.1	2.3319	—	—
J	B. C	25713	1807.7	2.3398	6.20	5.90
Ja	B. C	26038	1821.3	2.3694	1.75	1.65
Ja1	C. C''	26531	1842.1	2.4151	0.15	0.14
Ja2		27995	1905.2	2.5550	—	—
Ja3		28220	1915.2	2.5774	—	—
Jb	C''. Pb	28536	1929.1	2.6086	0.65	0.65
Jb1	C''. Pb	28812	1941.4	2.6364	0.15	0.15
Jb2		29168	1957.4	2.6726	0.15	0.15
Jb3	C. C''	29670	1980.2	2.7241	0.15	0.15
Jc	B. C	30657	2025.7	2.8280	0.75	0.75
Jc1	B. C	31835	2081.4	2.9566	0.15	0.15
Jc2	C. C''	36109	2296.6	3.4653	0.05	0.05
Jc3	C. C''	37599	2376.8	3.6599	0.05	0.05
Jc4	C. C''	39047	2457.4	3.8573	—	—
L	C''. Pb	41549	2603.1	4.2198	1.40	1.70
M	C''. Pb	46039	2886.6	4.9417	1.40	1.70
Ma	C''. Pb	46755	2934.6	5.0661	0.04	0.05
Ma1	C. C''	48192	3033.3	5.3229	—	—
N	C''. Pb	50019	3163.7	5.6658	0.25	0.30
Na	C''. Pb	50655	3210.4	5.7891	0.05	0.07
O		53341	3415.2	6.3352	0.20	0.30
O1		54788	3530.8	6.6461	0.02	0.03
O2		55220	3566.1	6.7415	0.02	0.03
O3		55559	3594.1	6.8175	0.01	0.01
Oa		55978	3629.0	6.9123	0.07	0.10
Oa1		56821	3700.1	7.1052	0.025	0.04
Oa2		57300	3741.1	7.2168	0.012	0.02
Oa3		58837	3875.9	7.5857	—	—
P		59377	3924.4	7.7163	0.20	0.30
Pa		60322	4010.7	7.9573	0.016	0.02
Pa1		62264	4194.1	8.4640	0.008	0.01
Pa2		62581	4224.8	8.5497	0.008	0.01
Pa3		63067	4272.4	8.6820	—	—
X	C''. Pb	00000	10000	25.31	0.11	0.35
Xa	C''. Pb	01115	10260	26.08	—	—
Xa1	C''. Pb	01284	10300	26.19	—	—

* The system of nomenclature in column I is that suggested in 'Nature,' vol. 129, p. 276 (1932).

† Entries are made in this column only where the origin seems fairly certain (see paragraphs 3 *et seq.*).

‡ The numerical values in columns III, IV, and V are given to more figures than is justified in order to facilitate checking the calculations.

§ The absence of a figure for the intensity indicates that the line was too weak to be measured by the photometer.

|| The accuracy of the figures for the intensities must be gauged from the description given of this part of the measurement. Below H ρ 1500 there is considerable uncertainty in the deduction of the true intensities from the photographic intensities.

This has some advantages, since the estimated accuracy to 1 part in 1000 represents an amount plus or minus of 0.00043 on each of each of these figures. When, from time to time, improvements in the absolute scale of these relative $H\beta$'s become necessary, it will be necessary only to add or subtract something from these logarithms. In this column, as in the others of Table II, more figures are given in each entry than are justified by the accuracy of the measurements. This is done, in view of the considerable amount of computation involved in handling these results, to avoid purely numerical errors arising from rounded figures and to permit cross-checking of the entries.

At first these relative values were translated into absolute values by taking the strong line F as standard with the value $H\beta$ 1398, but this began to lead to difficulties. It was found that the difference of energy between two lines which were known to be due to the same γ -ray, one arising from conversion in the K level and one from conversion in the L_I level, did not agree in giving the same value of $h\nu$ of the γ -ray when the respective absorption energies obtained from X-ray data were added to the energies of the β -ray lines. This point was most carefully checked, and many possible errors were looked into, but finally the conclusion was reached that either there was here a real effect or alternatively the whole scale of the absolute energies was too high.* This value for the standard line F of $H\beta$ 1398 was based on a direct comparison with the radium B values which I had determined about eight years ago.† The accuracy of these was estimated then to be 1 part in 500, whereas to avoid this discrepancy in the energies it would be necessary to lower the absolute values by nearly 1 per cent. It was therefore decided to repeat this measurement of the absolute values. Identically the same apparatus was used as had been used eight years ago, and lower values were at once obtained. It was very difficult to find any explanation of this, unless in the previous experiments there had been some magnetic disturbance from neighbouring workers which had not been properly taken into account. It was then clear that a completely new determination would have to be carried out. The magnetic field was measured by balancing the throw due to rotating a search coil through 180° against the throw of a Duddell inductometer. This latter instrument consists essentially of two fixed parallel coils through which a current can be sent and between which is mounted a third coil capable of rotation through 180° . The search coil and this Duddell rotating coil were coupled directly by a long brass rod so that they turned together. These two coils were then connected in series with

* Ellis, 'Nature,' vol. 129, p. 691 (1932).

† Ellis and Skinner, 'Proc. Roy. Soc.,' A, vol. 105, p. 165 (1924).

a fluxmeter, and the measurement consisted in adjusting the current in the primary of the Duddell until there was no throw on the fluxmeter when the coils turned over. The great advantage of this method is that the current in the Duddell remains steady and is easily measured accurately. It is not repeatedly made and broken as in the usual method of using a mutual inductance to balance against the search coil. The flux change from the Duddell is equal to the current flowing multiplied by the sum of its mutual inductance in its two positions. These were measured in each experiment by direct comparison with a mutual inductance. Both this mutual inductance and the Duddell were calibrated by the National Physical Laboratory. The chief source of error lies in the determination of the area turns of the search coil, and to minimise this three separate coils were used in these experiments, although for each the same marble core was used.

The line I (H_p 1751, energy 2.217×10^5 volts) was taken as the main reference line. Using the first search coil I obtained H_p values 1749 and 1752, and using the second 1751, giving a mean of 1751. Now using this value for I a comparison by ratios in two stages led to H_p values for the strong lines L and M of 2603 and 2887. A direct measurement using a third search coil gave 2608 and 2892, values differing only by 1 in 500.

The large permanent magnet used in these experiments greatly facilitated these measurements. The field could be measured at leisure both before and after the exposure, and there was no question of it changing during the photograph. It is greatly to be hoped that some other independent determination will be made soon of these absolute H_p values. It should be mentioned that these values agree well with an independent and early measurement of Meitner,* and in addition lead to energies which agree well with the X-ray data. There is some slight evidence that the values are still on the high side and may have to be lowered later by one or two parts in a thousand.

One absolute measurement was made for the very fast group X, and led to a value also rather more than $\frac{1}{2}$ per cent. below that previously accepted. The group of lines A-Bc have not yet been investigated, except as regards relative H_p 's. The H_p values are not on the same scale as the rest of the table and will possibly have to be lowered by a $\frac{1}{2}$ per cent.

Using this value for I, all the relative values of H_p have been transferred into absolute values as shown in column IV of Table II. From these the energies (column V) have been calculated. These have all been recalculated

*'Z. Physik,' vol. 11, p. 35 (1922).

for the present purpose, using a value for $e/m_0 = 1.760 \times 10^7$. Many new faint lines have been found in this work, but some faint lines reported by Black (*loc. cit.*) were not found.

It has already been mentioned that the main idea of this work was to measure the intensities of the groups; and this actually turned out to be far the most laborious part of the investigation. The majority of the lines only stand up but little from the continuous background formed by the disintegration electrons, and even careful photometering gives results which show considerable variations. The procedure was to take a series of photographs with the same radioactive source, to develop these together under carefully standardised conditions, and then to obtain the relative intensities of the photographic impressions of the lines in this group of plates. For convenience the main group of lines around I was usually taken as reference, and in each experiment one photograph of these lines was taken for comparison. To these experimental results two corrections have to be applied, first for what is termed the characteristic curve of the plate, which means the dependence of the blackening of the plate on the number of β -particles striking it. This was determined by means of a separate apparatus in which a plate could be exposed for various times to a constant source of β -particles. A calibration plate from this apparatus was included in each development batch. The validity of the reciprocity law has been well established for β -particles, and this calibration, using a constant source but different exposures, may therefore be used for comparing different numbers of electrons acting for the same time. Further, the intensities so corrected for the characteristic curve of the plate have then to be corrected for the different positions of the lines on the plate. Other things being equal, the greater the radius of curvature, the weaker is the effect of a given group. To a first approximation the intensity falls off inversely as ρ but a more accurate calculation has been made by Wooster* and his results were used for correcting the present measurements. His calculation was checked by photographing one group on the same plate but in different magnetic fields and noting the density of the impressions and the time of exposures.

It is difficult to give any definite figure for the accuracy of these results. Photometric work is extremely difficult on faint lines and in view of the number of lines which there were to photometer, it was impossible to take a sufficient number of measurements on each line to obtain the error experimentally. The values so obtained are shown in column VI of Table I under the heading

* 'Proc. Roy. Soc.,' A, vol. 114, p. 729 (1927).

"Photographic intensity," but this does not represent the relative number of electrons in the groups. To obtain these, it is necessary to know how the photographic effect of a given number of β -particles depends on their speed. This was investigated by Ellis and Aston* for β -particles of speeds 2000 $H\rho$ and upwards, covering only a small portion of the region studied in the present work.

An attempt to continue the photographic activity curve to lower values of $H\rho$ has been made in the following way. The curve so far as it is known from direct experiment is shown in full line in fig. 3. Now it is well known that the two lines F and I are due respectively to the conversion of a γ -ray in the K

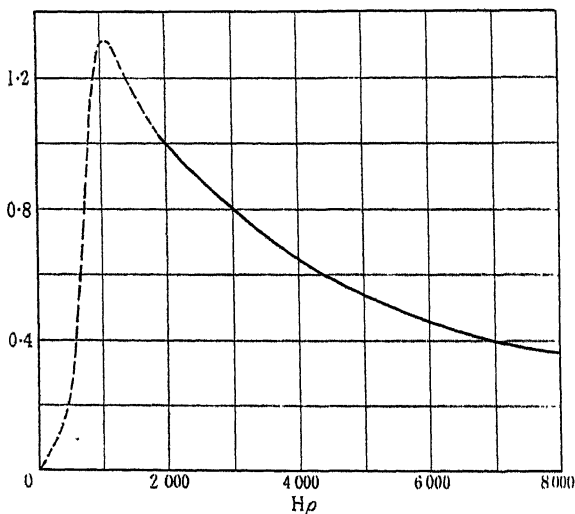


FIG. 3.

and L_I levels and there is some theoretical and experimental evidence that the relative intensities of the lines should be about 7 to 1. But the ratio of the photographic intensities, column VI, Table II, has been found to be 8 to 1, from which it would appear that, for the plates used, the photographic activity of electrons of $H\rho$ 1751 (line I) is about seven-eighths of those of $H\rho$ 1386 (line F). By extrapolation it is seen that the figure for $H\rho$ 1751 is 1.07, leading to a value of about 1.2 for $H\rho$ 1386. It is clear that the curve must pass through a maximum and tend towards zero at low speeds; a further point in this latter region can be obtained for some measurements of Gurney.† He measured the number of electrons in the strong lines by means of a Faraday

* 'Proc. Roy. Soc.,' A, vol. 119, p. 645 (1928).

† 'Proc. Roy. Soc.,' A, vol. 112, p. 380 (1926).

cylinder, and his results show that the relative intensity of F/A is about four or five to one. I have not yet made accurate measurements, but I estimate from some preliminary experiments that the photographic intensity of F relative to A is 20, from which it follows that the photographic activity of H ρ 543 is about 0.25 to 0.2.

These points suggest the continuation of the curve shown in dotted line in fig. 3. It is, of course, obvious that no great reliance can be placed on this curve, but it seems worth while putting it forward until such time as this region is studied experimentally.

The thickness of the sensitive layer of the photographic plates was about 0.0045 gm./cm.². Electrons of 750 H ρ could just penetrate this layer, but we would expect the photographic activity to increase for faster electrons since there will be a considerable reflection of electrons at the glass surface, and if these have enough energy they can again pass through the film. It is not therefore unreasonable to find the maximum for electrons of about 1000–1100 H ρ whose ranges according to Schonland* are respectively 0.0095 and 0.012 gm./cm.², that is, rather more than twice the thickness of the film.

This curve has been used to obtain the true intensities of the groups given in column VII of Table II, these figures show the relative numbers of electrons in the groups obtained from a source of thorium (B + C + C' + C'') in transient equilibrium.

§ 3. *The γ-rays of Th B . C.*

There are three γ-rays for which sufficient β-ray lines due to conversion in different levels are found to render their existence reasonably certain. The γ-rays are named after the most prominent line and the agreement of the energies and other relevant data is shown in Table III.

It will be noticed that in the case of γH the K line is about eight times as strong as the L_I which is again about five times as strong as the M_I line. This is in fair agreement with theory which predicts ratios of about 7 and 3.3 respectively. It must be remembered that we cannot adduce the evidence from the strongest γ-ray γF in support of this since the curve showing the dependence of photographic action on the speed of the β-particle was adjusted to give this result. It is interesting to note, however, that in this case the ratio of the L_I/M_I intensities is about 4.

We shall have occasion to make frequent use of these results, as for example in considering the γ-ray γE. It might be thought that the main β-ray line E

* 'Proc. Roy. Soc.,' A, vol. 104, p. 235 (1923) ; vol. 108, p. 187 (1925).

Table III.

β -ray line.	Intensity.	Level of conversion.	Energy volts $\times 10^{-5}$.	Absorption energy volts $\times 10^{-5}$.	$h\nu$ of γ -ray volts $\times 10^{-5}$.
γE E	2.30	L_I	0.9830	0.1635	1.1465
Ea	0.20	L_{II}	0.9892	0.1570	1.1462
Eb	0.75	M_I	1.1066	0.0401	1.1467
Eb1	0.20	N_I	1.1386	0.0096	1.1482
				Mean 1.147×10^5 volts	
γF F	165	K	1.4766	0.9000	2.3766
I	23.3	L_I	2.2170	0.1635	2.3805
Ia	1.5	L_{II}	2.2235	0.1570	2.3805
Ia1	0.08	L_{III}	2.2403	0.1340	2.3743
J	5.90	M_I	2.3398	0.0401	2.3799
Ja	1.65	N_I	2.3694	0.0096	2.3790
				Mean 2.379×10^5 volts.	
γH H	6.00	K	2.0892	0.9000	2.9892
Jc	0.75	L_I	2.8280	0.1635	2.9915
Jc1	0.15	M_I	2.9566	0.0401	2.9967
				Mean 2.990×10^5 volts.	

was due to conversion in the K level, but if this were so we should expect an L_I line of energy 1.720×10^5 volts and intensity 0.3. No such line has as yet been detected and I therefore prefer the allocation given in the table, which not only gives excellent agreement with the energies, but also shows the M_I line one-third the intensity of the L_I line. The β -ray line due to K conversion of this γ -ray should have an intensity of 16, and ought therefore to be easily detectable. It would, however, either coincide with or have slightly less energy than line A, and would be difficult to detect. The question whether it would coincide exactly or not cannot be answered at present as the lines A-Bc have not yet been compared with the main region.

There are several curious cases of exact superposition of lines, notably one in the thorium C'' . Pb spectrum which has already been referred to. In the Th B. C spectrum there may be another case involving the β -ray lines Dg, Fa and Ia2. For example :—

$$Dg + K_{abs} = 1.757 \quad \text{Intensity } 0.40$$

$$Fa + L_{Iabs} = 1.757 \quad \text{Intensity } 0.35.$$

While the energies are in excellent agreement Fa ought only to be of intensity about 0.06. But Fa may also be the K line to the extent of an intensity of 0.30 of another γ -ray and the corresponding L_I line does in point of fact exist, *i.e.*,

$$Fa + K_{abs} = 2.493.$$

$$Ia2 + L_{Iabs} = 2.495.$$

Its intensity would be only 0.04 which is in good agreement with it being so much weaker than Ia1 (intensity 0.08) that its intensity could not be measured. Against this view is that Black (*loc. cit.*) was of the opinion that Ia2 came from thorium C. Further investigation is necessary on this point, but provisionally the two γ -rays γ Dg and γ Fa will be assumed to come from thorium B . C.

There only remains one thorium B line, Ja3, to account for. This line is given by Black, but I was unable to convince myself that it really existed. It is, however, included in Table II in the relative position found by Black. These conclusions are collected in Table IV.

Table IV.—Thorium B . C γ -rays.

Name.	Energy in volts $\times 10^{-5}$.	Intensity of K line.
γ Dg	1.757	0.40
γ E	1.147	16
γ F	2.379	165
γ Fa	2.494	0.3
γ H	2.990	6.0

It will be noticed that

$$\gamma\text{Dg} - \gamma\text{E} = \gamma\text{H} - \gamma\text{F} = 0.611 \times 10^5 \text{ volts}$$

or alternatively

$$\gamma\text{H} + \gamma\text{E} = \gamma\text{Dg} + \gamma\text{F} = 4.136 \times 10^5 \text{ volts.}$$

There are two level systems which are suggested by this relation, which are shown in fig. 4.

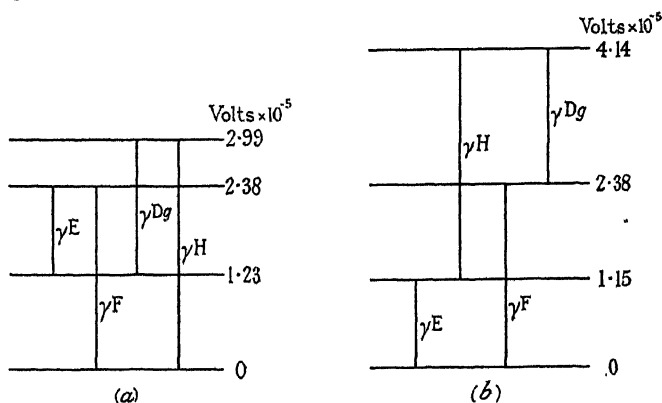


FIG. 4.—Possible levels of Th BC.

There is no definite evidence at present to decide which, if either, of these level systems is correct, but I think (b) is the more probable, since it fits in with the intensities quite plausibly.

§ 4. *The γ -rays of Thorium C . C''.*

The independent experimental evidence* for the association of γ -rays with this disintegration refers only to two γ -rays, γA which accounts for the β -ray lines A, Aa, Ab, B, Ba, Bb, Bc and γGa , giving the lines Ga and Jb3. Neither of these two γ -rays are suitable for an accurate test whether the energy differences of the β -ray lines agree with the X-ray data, since in the case of γA the absolute energy has not been determined, and is probably too high as given, and the β -ray lines due to γGa are rather too faint for accurate measurement. However, in this last case we find

$$\text{Intensities } \begin{cases} 0.65 \text{ Ga} + K = 2.0173 + 0.8515 = 2.869 \times 10^5 \\ 0.15 \text{ Jb3} + L_r = 2.7241 + 0.1532 = 2.877 \times 10^5. \end{cases}$$

With such a faint L_r line the absence of a better agreement either in the energies or in the ratio of the intensities, which should be about 7 to 1, can scarcely be considered significant.

The following provisional allocation of lines to thorium C . C'' is based on the assumption of the correctness of Gamow's† hypothesis connecting the fine structure of the α -ray groups with the γ -rays, and is carried out by searching for γ -rays which fit in with the α -particle measurements. Rosenblum and Valadres (*loc. cit.*) latest values for the energies of disintegration corresponding to the different α -particle groups are shown in Table V.

Table V.—Energies of Disintegration of Thorium C . C''.

Name of α -particle group.	Energy volts $\times 10^{-6}$.
α_1	6.189
α_0	6.148
α_2	5.857
α_4	5.712
α_3	5.691
α_5	5.563

The possible transitions that might be expected from this system are given in Table VI.

The γ -rays that are deduced from the α -ray measurements depend on the difference of two values, and if either were in error by 1 part in 1000 the resulting γ -ray would be in error by 0.06×10^5 volts. It will be noticed that an error

* Ellis, 'Proc. Roy. Soc.,' A, vol. 136, p. 396 (1932).

† 'Nature,' vol. 126, p. 396 (1930).

of this size is sufficient to account for the divergencies shown in the table. An error of only this amount implies an error in the actual measurement of the α -particle velocities not greater than 1 in 2000, and it is doubtful whether this accuracy has yet been obtained. We can only conclude that, while direct proof of the association of γ -rays with the α -particle energies is as yet unfeasible, by these means, the experimental results are entirely compatible with this view. A closer inspection of the last table lends weight to this conclusion. The level α_5 is associated with the weakest α -particle group, and it is scarcely surprising that only slender evidence is found for the corresponding γ -rays. The β -ray lines H1 and Ma1 were measured by Black but I did not find con-

Table VI.

Transition.	Energy calculated volts $\times 10^{-5}$.	γ -rays found volts $\times 10^{-5}$.
$\alpha_5 \rightarrow \alpha_3$	1.28
$\alpha_5 \rightarrow \alpha_4$	1.49
$\alpha_5 \rightarrow \alpha_2$	2.94	γ H1 2.98
$\alpha_5 \rightarrow \alpha_0$	5.85*
$\alpha_5 \rightarrow \alpha_1$	6.26	γ Ma1 6.17
$\alpha_3 \rightarrow \alpha_4$	0.21
$\alpha_3 \rightarrow \alpha_2$	1.66
$\alpha_3 \rightarrow \alpha_0$	4.57	γ Je3 4.51
$\alpha_3 \rightarrow \alpha_1$	4.98†
$\alpha_4 \rightarrow \alpha_3$	1.45
$\alpha_4 \rightarrow \alpha_0$	4.36	γ Je2 4.32
$\alpha_4 \rightarrow \alpha_1$	4.77	γ Je4 4.71
$\alpha_2 \rightarrow \alpha_0$	2.91	γ Ga 2.87
$\alpha_2 \rightarrow \alpha_1$	3.32	γ Ja1 3.27
$\alpha_0 \rightarrow \alpha_1$	0.41	γ A 0.40

* Even if present would be hidden by stronger Th C''. Pb β -ray Ma.

† Ditto, in this case by Th C''. Pb β -ray L.

clusive evidence for them on my plates and as yet have not made a further search for them. The remaining transitions which were not found may be either accounted for as indicated in notes for $\alpha_5 \rightarrow \alpha_0$ and $\alpha_3 \rightarrow \alpha_1$, or like $\alpha_5 \rightarrow \alpha_3$, $\alpha_5 \rightarrow \alpha_4$, and $\alpha_4 \rightarrow \alpha_2$, they fall in the X-ray region which has not yet been disentangled.

Considered as a whole, the agreement is very satisfactory and lends strong support to Gamow's hypothesis about the association of γ -rays and α -particle groups. I think the following are the most reasonable values that can be given at present for the energies of the excited states in excess of that of the ground state 0.41, 3.28, 4.73, 4.93, 6.26, in units of 10^5 volts.

§ 5. *The γ -rays of Th C'' . Pb.*

There are only four γ -rays which can with certainty* be associated with Th C'' . Pb. They are shown in Table VII.

Table VII.

β -ray line.	Intensity.	Level of conversion.	Energy volts $\times 10^{-5}$.	Absorption energy volts $\times 10^{-5}$.	$h\nu$ of γ -ray volts $\times 10^{-5}$.
γ G G	3.20	K	1.8868	0.8750	2.7618
Jb	0.65	L _I	2.6086	0.1582	2.7668
Jb1	0.15	L _{III}	2.6364	0.1301	2.7665
				Mean 2.765 $\times 10^5$ volts.	
γ L L	1.70	K	4.2198	0.8750	5.0948
M	1.70	L _I	4.9417	0.1582	5.0999
Ma	0.05	M _I	5.0661	0.0380	5.1041
				Mean 5.100 $\times 10^5$ volts.	
γ M M	1.70	K	4.9417	0.8750	5.8167
N	0.30	L _I	5.6658	0.1582	5.8240
Na	0.07	M _I	5.7891	0.0385	5.8276
				Mean 5.823 $\times 10^5$ volts.	
γ X X	0.35	K	25.31	0.8750	26.19
Xa	—	L _I	26.08	0.1582	26.24
Xa1	—	M _I	26.19	0.0385	26.23
				Mean 26.20 $\times 10^5$ volts.	

Doubtless some of the remaining faint lines, of which there are a considerable number, are due to γ -rays emitted from Th C'' . Pb, but there is at present no evidence to justify any attempt to sort these out.

The γ -rays of Th C'' . Pb.

Name.	Energy in volts $\times 10^{-5}$.	Intensity of K line.
γ G	2.765	3.2
γ L	5.100	1.7
γ M	5.823	1.5
γ X	26.20	0.35

§ 6. *The γ -rays of Thorium C . C'.*

Thorium C' emits several groups of long-range α -particles, indicating several different states of excitation of this nucleus previous to disintegration. We should therefore expect γ -rays to be emitted as an alternative process to the

* Ellis, *loc. cit.*

emission of long-range α -particles, and the $h\nu$ of these γ -rays should be equal to the difference in energy of the long-range and normal α -particles. Such γ -rays would, in the terminology adopted here, be due to thorium C. C'. Rosenblum and Valadres (*loc. cit.*) values for the energies of the α -particle groups are shown in Table VIII and also the corresponding energy of disintegration taking into account the recoil. The normal disintegration gives the group α_0 , and the long-range groups in order of intensity are α_1 , α_{11} , α_{111} . The transition that one might expect from these excited states are also shown.

Table VIII.— α -ray Groups of Thorium C'.

Name.	Energy of group volts $\times 10^{-3}$.	Energy of disintegration volts $\times 10^{-3}$.
α_1	10.52	10.72
α_{111}	9.60	9.78
α_{11}	9.45	9.63
α_0	8.77	8.94

Table IX.—Possible Transitions.

Transition.	Energy volts $\times 10^{-5}$.
$\alpha_1 - \alpha_0$	17.8
$\alpha_{111} - \alpha_0$	8.4
$\alpha_{11} - \alpha_0$	6.9
$\alpha_1 - \alpha_{11}$	10.9
$\alpha_{111} - \alpha_{11}$	1.5
$\alpha_1 - \alpha_{111}$	9.4

In agreement with Black I could find no strong β -ray lines above an energy of 8.7×10^5 volts except those due to the thorium C". Pb γ -ray at 25×10^5 volts. In some photographs I obtained indications of faint lines but I would like to investigate the matter further before making any definite statement. It seems quite likely, however, that γ -rays corresponding to these transmissions are actually emitted, and in fact Skobeltzyn (*loc. cit.*) by his method of detection using the Compton effect has noticed a relatively powerful emission in the region around 17×10^5 volts. A possible conclusion is that we have here abnormally low internal conversion.

There are a great many β -ray lines between 6.3×10^5 and 8.7×10^5 , some of which may possibly be associated with the above states of excitation. For example the lines O and Oa1 fit well both in energy and intensity with a γ -ray of about 7.26×10^5 volts converted in atomic number 84.

Table X.

Name.	Intensity.	Energy volts $\times 10^{-5}$.	Absorption energy volts $\times 10^{-6}$.	$h\nu$ of γ -ray volts $\times 10^{-5}$.
O	0.30	6.335	0.926	7.26
Oa1	0.04	7.105	0.169	7.27

If this is to correspond to the transition $\alpha_{11}-\alpha_0$, then the α -ray group α_{11} would have to have an energy of 9.49×10^6 volts instead of 9.45×10^6 as measured. Taking into account the difficulty of the α -particle measurement, an error of this amount is perhaps not impossible, but it is clearly out of the question to come to any definite decision. In a similar manner the β -ray lines P and Pa1 suggest a γ -ray of energy 8.64×10^5 volts, agreeing not badly with the transitions $\alpha_{111}-\alpha_0$. The difficulty here is that the α -ray group α_{111} is so weak that it is rather surprising to find the β -ray lines.

Considered as a whole it seems possible that there is a discrepancy between the β -ray lines one would expect and those that are found. As has already been said, however, the explanation may lie partly in slight errors in the α -particle measurements and partly that we have in this body abnormally low internal conversion.

§ 7. *The γ -rays of Th C' . Pb.*

The absence of any fine structure in the main group of α -particles from Th C' suggests that the nucleus is not left excited, and that therefore there are no γ -rays emitted by Th C' . Pb.

It is a pleasure to thank Lord Rutherford for his continued interest and help in these measurements, and also the Government Grant Committee of the Royal Society for grants with which radioactive material and the large permanent magnet were purchased. Professor Schlundt of the University of Missouri very kindly gave me some radio-thorium which was used in the later experiments, and for which I am glad to take this opportunity of thanking him. I would also like to acknowledge the assistance I received from Mr. R. Cole.

Summary.

The β -ray spectrum of thorium (B + C) has been investigated using the semicircular focussing method and photographic registration. New measurements have been made of the energies of the β -ray groups. These consisted,

in the first place of a determination of the relative $H\rho$'s of the different groups and in the second of a measurement in absolute units of the $H\rho$'s of certain selected strong groups. The values obtained are about 0·7 per cent. lower than those which has been previously accepted. The intensities of the groups have also been measured photometrically. An attempt is made to deduce from these results the γ -rays associated with the various disintegrating bodies.

The Theory of Wave Resistance.

By T. H. HAVELOCK, F.R.S.

(Received August 6, 1932.)

Introduction.

1. In the following paper general expressions are obtained for the wave resistance of a continuous distribution of sources and sinks over a surface within the liquid, and also for a similar distribution of normal doublets. These expressions follow directly from results given previously,* and may be applied to give the wave resistance of any solid for which a suitable distribution of sources or doublets over its surface can be found.

The opportunity is taken to give, for comparison, the similar results for a distribution of pressure over the surface of the liquid, using the same notation and the same general method of calculating the wave resistance.

The various results are discussed briefly in relation to the ship problem. Certain interpolation formulæ, of a semi-empirical nature, have been proposed recently in attempting to extend the range of existing expressions for the wave resistance of a ship; these are shown to have their interpretation as particular cases of source distributions of the nature considered here.

Source Distribution.

2. We begin with a simple point source of strength m at a depth f below the free surface of the liquid, and suppose the source to be moving horizontally in the direction Ox with uniform velocity c . Take the origin O in the free surface with Oz vertically upwards, the source being at the point $(0, 0, -f)$ referred to

* 'Proc. Roy. Soc.,' A, vol. 118, p. 24 (1928).

moving axes. Let ζ be the surface elevation, and assume a frictional force in the liquid proportional to velocity, the frictional coefficient being ultimately made zero.

The pressure condition at the free surface is

$$\frac{\partial \phi}{\partial t} - g\zeta + \mu' \phi = \text{constant}, \quad (1)$$

and this gives

$$\frac{\partial^2 \phi}{\partial x^2} + \kappa_0 \frac{\partial \phi}{\partial z} - \mu \frac{\partial \phi}{\partial x} = 0, \quad (2)$$

at $z = 0$, with $\kappa_0 = g/c^2$ and $\mu = \mu'/c$. Assume for the velocity potential

$$\begin{aligned} \phi = \frac{m}{2\pi} \int_{-\pi}^{\pi} d\theta \int_0^{\infty} e^{-\kappa(z+f) + i\kappa\varpi} d\kappa \\ + \int_{-\pi}^{\pi} d\theta \int_0^{\infty} F(\kappa, \theta) e^{\kappa z + i\kappa\varpi} d\kappa, \end{aligned} \quad (3)$$

where $\varpi = x \cos \theta + y \sin \theta$, and the real part of the expression is to be taken. The first term in (3) gives the velocity potential of the given source, namely m/r_1 , in a form valid for $z + f > 0$. From the surface condition (2) we obtain

$$F(\kappa, \theta) = - \frac{m \kappa + \kappa_0 \sec^2 \theta + i\mu \sec \theta}{2\pi \kappa - \kappa_0 \sec^2 \theta + i\mu \sec \theta} e^{-\kappa f}. \quad (4)$$

Hence we may write the solution in the form

$$\phi = \frac{m}{r_1} - \frac{m}{r_2} - \frac{\kappa_0 m}{\pi} \int_{-\pi}^{\pi} \sec^2 \theta d\theta \int_0^{\infty} \frac{e^{-\kappa(f-z) + i\kappa\varpi}}{\kappa - \kappa_0 \sec^2 \theta + i\mu \sec \theta} d\kappa, \quad (5)$$

where

$$r_1^2 = x^2 + y^2 + (z + f)^2; \quad r_2^2 = x^2 + y^2 + (z - f)^2.$$

It is to be understood that the limiting value of (5) is taken for $\mu \rightarrow 0$.

We may now generalise by integration. We replace x and y by $x - h$ and $y - k$ respectively, and take σ to be the surface density of source at a point $(h, k, -f)$ on a surface S within the liquid. Thus the velocity potential is given by

$$\begin{aligned} \phi = \int \left(\frac{1}{r_1} - \frac{1}{r_2} \right) \sigma dS \\ - \frac{\kappa_0}{\pi} \int \sigma dS \int_{-\pi}^{\pi} \sec^2 \theta d\theta \int_0^{\infty} \frac{e^{-\kappa(f-z) + i\kappa\varpi}}{\kappa - \kappa_0 \sec^2 \theta + i\mu \sec \theta} d\kappa, \end{aligned} \quad (6)$$

with

$$r_1^2 = (x - h)^2 + (y - k)^2 + (z + f)^2$$

$$r_2^2 = (x - h)^2 + (y - k)^2 + (z - f)^2$$

$$\varpi = (x - h) \cos \theta + (y - k) \sin \theta.$$

It is assumed that the distribution is such that the various integrals are convergent.

3. To calculate the wave resistance R we use the method of the previous paper to which reference has already been made. With the inclusion of the frictional term in the equations of fluid motion, energy is dissipated at a rate equal to $2\mu'$ times the total kinetic energy of the liquid and this must be equal to the product Rc . As μ' is made to approach zero the quantity so calculated approaches a finite limiting value, and its physical interpretation in the limit when there is no fluid friction is the rate at which energy is propagated outwards in the wave motion.

The rate of dissipation of energy is given by

$$- \mu' \rho \int \phi \frac{\partial \phi}{\partial n} dS, \quad (7)$$

taken over the boundaries of the liquid. As we require only the limiting value, we have the wave resistance given by

$$R = \lim_{\mu \rightarrow 0} \mu \rho \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \phi \frac{\partial \phi}{\partial z} dx dy, \quad (8)$$

taken over the free surface $z = 0$.

Referring to (6), and putting the first two terms in the same integral form as the third, we obtain, at $z = 0$,

$$\phi = -\frac{\kappa_0}{\pi} \int \sigma dS \int_{-\pi}^{\pi} \sec^2 \theta d\theta \int_0^{\infty} \frac{e^{-\kappa f + i\kappa \varpi}}{\kappa - \kappa_0 \sec^2 \theta + i\mu \sec \theta} d\kappa, \quad (9)$$

$$\frac{\partial \phi}{\partial z} = -\frac{1}{\pi} \int \sigma dS \int_{-\pi}^{\pi} d\theta \int_0^{\infty} \frac{\kappa^2 e^{-\kappa f + i\kappa \varpi}}{\kappa - \kappa_0 \sec^2 \theta + i\mu \sec \theta} d\kappa, \quad (10)$$

where the real parts are to be taken.

After some reduction, we may write the real part of (9) in the form

$$\begin{aligned} \phi = & \int_{-\pi}^{\pi} d\theta \int_0^{\infty} \{F_1 \cos(\kappa x \cos \theta) \cos(\kappa y \sin \theta) \\ & + F_2 \sin(\kappa x \cos \theta) \cos(\kappa y \sin \theta) + F_3 \cos(\kappa x \cos \theta) \sin(\kappa y \sin \theta) \\ & + F_4 \sin(\kappa x \cos \theta) \sin(\kappa y \sin \theta)\} \kappa d\kappa, \end{aligned} \quad (11)$$

in which

$$\begin{aligned} F_1 &= -\{(\kappa - \kappa_0 \sec^2 \theta) P_s - \mu Q_s \sec \theta\} D \\ F_2 &= -\{(\kappa - \kappa_0 \sec^2 \theta) Q_s + \mu P_s \sec \theta\} D \\ F_3 &= -\{(\kappa - \kappa_0 \sec^2 \theta) Q_0 - \mu P_0 \sec \theta\} D \\ F_4 &= -\{(\kappa - \kappa_0 \sec^2 \theta) P_0 + \mu Q_0 \sec \theta\} D \\ D &= \kappa_0 \sec^2 \theta / \pi \kappa \{(\kappa - \kappa_0 \sec^2 \theta)^2 + \mu^2 \sec^2 \theta\}, \end{aligned} \quad (12)$$

and the quantities P , Q are given in terms of the source distribution by

$$\begin{aligned} P_s &= \int \sigma e^{-\pi f} \cos(\kappa h \cos \theta) \cos(\kappa k \sin \theta) dS \\ P_0 &= \int \sigma e^{-\kappa f} \sin(\kappa h \cos \theta) \sin(\kappa k \sin \theta) dS \\ Q_s &= \int \sigma e^{-\kappa f} \sin(\kappa h \cos \theta) \cos(\kappa k \sin \theta) dS \\ Q_0 &= \int \sigma e^{-\kappa f} \cos(\kappa h \cos \theta) \sin(\kappa k \sin \theta) dS. \end{aligned} \quad (13)$$

Similarly from (10), $\partial\phi/\partial z$ is obtained in the same integral form as in (11), with quantities G instead of F given by the same expressions as in (12) but with

$$D = \kappa / \pi \{(\kappa - \kappa_0 \sec^2 \theta)^2 + \mu^2 \sec^2 \theta\}.$$

The expressions for the surface values of ϕ and $\partial\phi/\partial z$ are now in a form to which we may apply a theorem derived from the Fourier integral theorem in two variables; namely, we have, with the above notation

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \phi \frac{\partial\phi}{\partial z} dx dy = 4\pi^2 \int_{-\pi}^{\pi} d\theta \int_0^{\infty} (F_1 G_1 + F_2 G_2 + F_3 G_3 + F_4 G_4) \kappa d\kappa. \quad (14)$$

Using (8), this reduces readily to

$$\begin{aligned} R &= \lim_{\mu \rightarrow 0} 4\kappa_0 \rho \mu \int_{-\pi}^{\pi} \sec^2 \theta d\theta \int_0^{\infty} \frac{\kappa (P_s^2 + P_0^2 + Q_s^2 + Q_0^2)}{(\kappa - \kappa_0 \sec^2 \theta)^2 + \mu^2 \sec^2 \theta} d\kappa \\ &= 16\pi \kappa_0^2 \rho \int_0^{\frac{1}{2}\pi} (P_s^2 + P_0^2 + Q_s^2 + Q_0^2) \sec^3 \theta d\theta, \end{aligned} \quad (15)$$

where in (15) the quantities P and Q have the values given by (13) when κ has been replaced by $\kappa_0 \sec^2 \theta$.

This result may also be put in the form

$$R = 8\pi\kappa_0^2\rho \int_{-\frac{1}{2}\pi}^{\frac{1}{2}\pi} (P^2 + Q^2) \sec^3 \theta \, d\theta, \quad (16)$$

with

$$\left. \begin{matrix} P \\ Q \end{matrix} \right\} = \int \sigma e^{\kappa_0 z \sec^3 \theta} \frac{\cos}{\sin} \{ \kappa_0 (x \cos \theta + y \sin \theta) \sec^2 \theta \} \, dS. \quad (17)$$

In (17), the co-ordinates $(h, k, -f)$ have been replaced by current co-ordinates (x, y, z) ; since the sources are within the liquid, z is negative over the surface S .

Doublet Distribution.

4. A surface distribution of normal doublets could be obtained by generalising an expression for any two doublets, but it can be deduced directly from (16) and (17). We have simply to regard the surface S in (17) as a double sheet with source densities σ and $-\sigma$ respectively, and then proceed to the limit in the usual manner. The required result is obtained by applying the operator

$$l \frac{\partial}{\partial x} + m \frac{\partial}{\partial y} + n \frac{\partial}{\partial z},$$

to the expressions in (17), (l, m, n) being the direction of the normal to the surface. If M is the doublet moment per unit area, the axes being everywhere normal to the surface S , we obtain, in this way, the wave resistance

$$R = 8\pi\kappa_0^4\rho \int_{-\frac{1}{2}\pi}^{\frac{1}{2}\pi} (P^2 + Q^2) \sec^7 \theta \, d\theta, \quad (18)$$

in which

$$\begin{aligned} P &= \int M e^{\kappa_0 z \sec^3 \theta} \{ -(l \cos \theta + m \sin \theta) \sin (\kappa_0 \varpi \sec^2 \theta) \\ &\quad + n \cos (\kappa_0 \varpi \sec^2 \theta) \} \, dS \\ Q &= \int M e^{\kappa_0 z \sec^3 \theta} \{ (l \cos \theta + m \sin \theta) \cos (\kappa_0 \varpi \sec^2 \theta) \\ &\quad + n \sin (\kappa_0 \varpi \sec^2 \theta) \} \, dS, \end{aligned} \quad (19)$$

with $\varpi = x \cos \theta + y \sin \theta$.

These expressions may be put into various alternative forms, and, of course, may be simplified when the surface distribution is symmetrical with respect to the co-ordinate planes. It may be remarked that an expression given previously for the wave resistance of any two finite doublets in given positions may be deduced as a particular case of these results.

Pressure Distribution.

5. The wave resistance for a travelling distribution of pressure applied to the upper surface of the liquid has been worked out by various methods, but not by that used in the previous sections. It is convenient, for comparison, to have the general case set out in the same way and using the same principle for the calculation of the resistance.

We begin by assuming a possible form for the velocity potential and finding the surface pressure to which it corresponds.

We take

$$\phi = \int_{-\pi}^{\pi} \sec \theta \, d\theta \int_0^{\infty} \frac{F(\kappa) e^{\kappa z + i\kappa w}}{\kappa - \kappa_0 \sec^2 \theta + i\mu \sec \theta} \kappa \, d\kappa, \quad (20)$$

with $w = x \cos \theta + y \sin \theta$.

From the kinematical condition at $z = 0$, the surface elevation is given by

$$\zeta = -\frac{i}{c} \int_{-\pi}^{\pi} \sec^2 \theta \, d\theta \int_0^{\infty} \frac{F(\kappa) e^{i\kappa w}}{\kappa - \kappa_0 \sec^2 \theta + i\mu \sec \theta} \kappa \, d\kappa. \quad (21)$$

The pressure at the surface ($z = 0$) is found from

$$\frac{p}{\rho} = -c \frac{\partial \phi}{\partial x} - g\zeta + \mu' \phi. \quad (22)$$

Using (20) and (21), this reduces to

$$\begin{aligned} p &= -ic\rho \int_{-\pi}^{\pi} d\theta \int_0^{\infty} \kappa F(\kappa) e^{i\kappa(x \cos \theta + y \sin \theta)} d\kappa \\ &= -2\pi\rho ci \int_0^{\infty} \kappa F(\kappa) J_0(\kappa r) d\kappa, \end{aligned} \quad (23)$$

where $r^2 = x^2 + y^2$. Since we may write

$$p(r) = \int_0^{\infty} J_0(\kappa r) \kappa \, d\kappa \int_0^{\infty} p(\alpha) J_0(\kappa \alpha) \alpha \, d\alpha, \quad (24)$$

we see that

$$\phi = \frac{i}{2\pi c\rho} \int_{-\pi}^{\pi} \sec \theta \, d\theta \int_0^{\infty} \frac{f(\kappa) e^{\kappa z + i\kappa w}}{\kappa - \kappa_0 \sec^2 \theta + i\mu \sec \theta} \kappa \, d\kappa, \quad (25)$$

represents the solution for a surface pressure $p(r)$, symmetrical round the moving origin, with

$$f(\kappa) = \int_0^{\infty} p(\alpha) J_0(\kappa \alpha) \alpha \, d\alpha. \quad (26)$$

To generalise this, we first suppose the pressure concentrated round the origin and of integrated amount P , so that $f(\kappa)$ in (25) is replaced by $P/2\pi$. Then for

any continuous distribution of pressure $p(x, y)$, we obtain by integration

$$\phi = \frac{i}{4\pi^2 c \rho} \int p(h, k) dS \int_{-\pi}^{\pi} \sec \theta d\theta \int_0^{\infty} \frac{e^{\kappa z + i\kappa \varpi}}{\kappa - \kappa_0 \sec^2 \theta + i\mu \sec \theta} \kappa d\kappa, \quad (27)$$

where now we have $\varpi = (x - h) \cos \theta + (y - k) \sin \theta$.

6. We obtain the corresponding wave resistance from the rate of dissipation of energy exactly as in the previous sections, and we use the formula (8). The surface values of ϕ and $\partial\phi/\partial z$ are put into the form (11) and the calculation carried out as in (14). From the similarity of the forms for ϕ in the two cases, the result may be written down. We obtain

$$\begin{aligned} R &= \lim_{\mu \rightarrow 0} \frac{\mu}{4\pi^2 c^2 \rho} \int_{-\pi}^{\pi} \sec^2 \theta d\theta \int_0^{\infty} \frac{\kappa^2 (P_e^2 + P_0^2 + Q_e^2 + Q_0^2)}{(\kappa - \kappa_0 \sec^2 \theta)^2 + \mu^2 \sec^2 \theta} d\kappa \\ &= \frac{\kappa_0^2}{\pi c^2 \rho} \int_0^{\frac{1}{2}\pi} (P_e^2 + P_0^2 + Q_e^2 + Q_0^2) \sec^5 \theta d\theta, \end{aligned} \quad (28)$$

where the quantities P and Q are as in (13) with f zero and σ replaced by p .

We may also write this in the form

$$R = \frac{\kappa_0^2}{2\pi c^2 \rho} \int_{-\frac{1}{2}\pi}^{\frac{1}{2}\pi} (P^2 + Q^2) \sec^5 \theta d\theta, \quad (29)$$

with

$$\left. \begin{matrix} P \\ Q \end{matrix} \right\} = \int p(x, y) \frac{\cos}{\sin} \{ \kappa_0 (x \cos \theta + y \sin \theta) \sec^2 \theta \} dS, \quad (30)$$

the latter integrations extending over the given surface distribution of pressure.

We may obtain an alternative form by integrating with respect to x in (30); provided the pressure distribution is continuous and is zero at its outer boundaries, we then have

$$R = \frac{1}{2\pi c^2 \rho} \int_{-\frac{1}{2}\pi}^{\frac{1}{2}\pi} (P^2 + Q^2) \sec^3 \theta d\theta, \quad (31)$$

with

$$\left. \begin{matrix} P \\ Q \end{matrix} \right\} = \int \frac{\partial p}{\partial x} \frac{\cos}{\sin} \{ \kappa_0 (x \cos \theta + y \sin \theta) \sec^2 \theta \} dS. \quad (32)$$

We may compare (31) and (32) with the expressions (16) and (17) for a distribution of sources on a surface within the liquid. Suppose we may neglect the depth of this latter surface at every point; then without considering the actual surface elevation, which would require a closer examination, we may say that the wave resistance for the two cases would be the same with the connection between the source density and the pressure distribution given by $4\pi g \rho \sigma = c \partial p / \partial x$.

Moving Solid.

7. An obvious application of these results is to the uniform motion of a submerged solid when we replace the solid by a distribution of sources or doublets over its surface; for a first approximation we may take the distribution to be that appropriate to the motion of the solid in an infinite liquid. This will, of course, give the same result as if we had used the system of sources and sinks which is the image of a uniform stream in the solid, or, in fact, any equivalent surface or volume distribution on or within the surface of the solid. Simple forms, such as the sphere or ellipsoid, for which the wave resistance has already been found, have been calculated from the known image system. For instance, the sphere was replaced by a doublet at the centre; it can be verified, after some reduction of integrals, that the expressions (16) and (17) with the proper value of σ over the surface of the sphere, lead to the same result for the wave resistance. In general, the expressions (16) and (17) allow the wave resistance to be calculated for solids for which an image system is not known, but for which the distribution of surface density can be determined by known methods of approximation.

Consider now an open plane distribution of sources and sinks over the vertical zx -plane. In this case the normal fluid velocity at a point on either side is $2\pi\sigma$, where σ is the source density at the point. For a ship of slender form, and small beam, symmetrical about the zx -plane, the normal velocity is taken to be approximately $c \partial y / \partial x$ if the surface of the ship is given by an equation $y = f(z, x)$. From (16) and (17), the usual expression for the wave resistance follows:

$$R = \frac{2g^2\rho}{\pi c^2} \int_{-\frac{1}{2}\pi}^{\frac{1}{2}\pi} (P^2 + Q^2) \sec^3 \theta \, d\theta, \quad (33)$$

$$\left. \begin{matrix} P \\ Q \end{matrix} \right\} = \iint \frac{\partial y}{\partial x} e^{\kappa_0 z \sec \theta} \frac{\cos}{\sin} (\kappa_0 x \sec \theta) \, dx \, dz, \quad (34)$$

the latter integrations being taken over the vertical longitudinal section.

For the other extreme case, a ship of flat form and small draught, comparison is usually made with a suitable distribution of pressure applied to the surface of the water, with the wave resistance given by, say, (31) and (32).

The similarity between the expressions for the resistance in these two extreme forms has been remarked upon by Weinblum,* and more recently by Hogner.† In an attempt to cover both cases by a single expression, Hogner has proposed

* C. Weinblum, 'Z. Math. Mech.,' vol. 10, p. 458 (1930).

† E. Hogner, 'Jahrb. Schiffbautech. Ges.' (1932).

a so-called interpolation formula which, when put into the notation of the present paper, is

$$R = \frac{g^2 \rho}{2\pi c^2} \int_{-\frac{1}{2}\pi}^{\frac{1}{2}\pi} (P^2 + Q^2) \sec^3 \theta \, d\theta, \quad (35)$$

$$\left. \begin{matrix} P \\ Q \end{matrix} \right\} = \iint \frac{\partial z}{\partial x} e^{\kappa_0 z \sec^3 \theta} \frac{\cos}{\sin} \{ \kappa_0 (x \cos \theta + y \sin \theta) \sec^2 \theta \} \, dx \, dy. \quad (36)$$

In (36) the integrations are taken over the section of the ship by the water surface, and the surface of the ship is given by an equation $z = F(x, y)$. It may be noted that if dS_y and dS_x are the projections of an element of the surface upon the zx -plane and the xy -plane respectively, we have

$$(\partial z / \partial x) dS_x = (\partial y / \partial x) dS_y.$$

In the limit $y \rightarrow 0$, (36) becomes equivalent to (34) under the conditions for a ship of small beam. On the other hand, in the limit $z \rightarrow 0$, (36) reduces to the expression (31) for a pressure distribution with the assumption $p = g\rho\zeta$. Without discussing this argument, it may be remarked that (36) is a particular case of the expressions in (16) and (17) for a distribution of sources over the surface of the ship. In the one extreme case, the narrow ship, we take $\sigma = (c/2\pi) \partial y / \partial x$, the sources forming in the limit a plane distribution. For the other extreme, the flat ship, a similar approximation would be $\sigma = (c/2\pi) \partial z / \partial x$. But it is only in these cases, when the source distribution approximates to a plane, that the normal fluid velocity can be expressed simply in terms of the source density; these expressions do not hold when the distribution is on a curved surface or, in other words, when the finite beam of the ship is taken into account.

It has been remarked that formulæ in use at present are in effect special cases of the general expressions (16) and (17), with simple approximations to the density of the source distribution. If we think of the distribution, appropriate to motion in an infinite liquid, as a suitable first approximation, it might be suggested that this should be used over the curved surface of the ship instead of the present simple expressions over the vertical longitudinal plane. In one sense this would be an improvement, but it is not likely that it would give any better agreement with experimental results; for the more we depart from the simple narrow ship the more necessary it is to take into account the effect of the wave motion upon the distribution of fluid velocity round the ship.

Instead of attempting to assign in advance a distribution of sources or doublets over the surface of the ship, it might be left to be determined, from

suitable integral equations, so that all the conditions of the problem should be satisfied. This, in itself, would not amount to more than a formulation of the general problem in different terms and would not advance its practical solution, unless possibly such a form of statement should lead to improved methods of approximation for the equivalent distribution.

The Resistivity of Polycrystalline Wires in Relation to Plastic Deformation, and the Mechanism of Plastic Flow.

By E. N. DA C. ANDRADE, Quain Professor of Physics in the University of London, and B. CHALMERS, Ph.D.

(Communicated by L. N. G. Filon, F.R.S.—Received February 12, 1932—
Revised August 9, 1932.)

1. *Introduction.*

The subject of metallic conduction of electricity has long been a source of difficulties, which are far from having been overcome by the many admirable and ingenious theories recently put forward.* The peculiarities of the different individual metals, as exemplified, for instance, in the variation of the Wiedemann-Franz ratio from metal to metal seem to fall right outside any scheme so far elaborated. At present it is only possible to treat a metal as if it were a homogeneous crystal, the insufficiency of which assumption may be the source of many of those discrepancies between theory and experiment now attributed to fundamental properties of the electron in metal. Any experimental evidence concerning variations in electrical conductivity which can be produced in one pure metal by mechanical treatment should therefore be of some interest for the general problem. In particular the effect of permanent deformation upon electrical conductivity in metals is full of obscurity,† and seems to offer a field for experiment.

It was decided to investigate the variations of conductivity with certain simple types of deformation which attend the flow of soft metals. The flow of polycrystalline wires under constant stress has been investigated by one of us,‡

* See, e.g., Hume-Rothery, "The Metallic State," (1931), p. 291, *et seq.*

† See, e.g., Hume-Rothery, *loc. cit.*, p. 27.

‡ Andrade, 'Proc. Roy. Soc.,' A, vol. 84, p. 1 (1910), and A, vol. 90, p. 392 (1914).

and the changes of mechanical properties with time examined. The process of flow can be analysed into three phases:—an initial stretch which occurs within a very short time of the application of the stress; a continuous flow, the rate of which decreases with time, called, from the constant used to express it, the β -flow; and a flow at constant rate per unit length which occurs simultaneously with the β -flow and continues until the wire breaks. The formula expressing the flow is

$$l = l_0(1 + \beta t^{1/3}) e^{kt}, \quad (1)$$

where l is the length at time t , and l_0 , β and k are constants. Since the formula was first put forward* analogous formulæ, involving $t^{1/3}$, have been found by other workers for different substances. Thus Filon and Jessop found for celluloid† $l = l_0 + at^{1/3} + bt$, and Peirce‡ has expressed the decrease of the couple required to maintain constant torsion in a cotton fibre by a formula $c = c_0 + ae^{-\beta t^{1/3}}$, so that a term in $t^{1/3}$ apparently represents with some generality the observed behaviour of solids during flow. In former publications the general features of the behaviour were expressed in terms of crystal structure,§ but at the time no evidence from other sources could be adduced to support the hypothesis. The object of the present investigation is to measure the specific resistivity of the metal during the plastic extension, and to use the data so obtained to elucidate the changes of structure which take place in a polycrystalline wire when it is extended. One of the chief results has been to show that the β -flow has a real physical significance, and is due to the rotation of the axes of the crystallites which constitute the polycrystalline wire. Experiments have also been carried out on the resistivity of single crystal wires, the results of which will be published in the near future.

To compare the specific resistivity of the wire at any stage of the extension with that of the unstretched wire, it was necessary to measure simultaneously the length and the resistance of the wire flowing under large constant stress. The metal first investigated was cadmium, in the form of wire, of diameter 0.046 cm., which was annealed for 6 hours at 100° before use. It was established by trial that higher temperature or longer time of annealing had no further effect on the properties of the wire that were being investigated. Experiments were also carried out with wires of copper, aluminium and tin,

* Andrade, *loc. cit.* (1910).

† 'Phil. Trans.,' A, vol. 223, p. 89 (1922).

‡ 'Shirley Institute Memoirs,' vol. 2, p. 278 (1923).

§ Andrade, *loc. cit.*

in all of which cases the annealing was performed under the same conditions as for cadmium.

2. *Apparatus for the Application of the Stress.*

An apparatus was designed with the following objects :—

- (i) To apply a constant stress to a wire kept at a constant temperature.
- (ii) To measure the length of the wire, and
- (iii) To measure the resistance of the wire.

One end of the wire, of length 20 cm., was fixed centrally in the bottom of a brass tube and a vertical upward pull was applied to the other end of the wire, by a method to be described later. The brass tube was fixed vertically at its upper end to a horizontal wooden board that carried both the optical and mechanical systems. The tube was immersed in a thermostatic bath below the board, and was pierced with holes to allow the water of the bath to circulate inside it. As the water of the thermostat bath was found to have some action on the wire, causing the resistance to rise, even when no extension was taking place, a glass tube, filled with oil, was placed inside the brass tube, the clamp which held the lower end of the wire passing through a cork in the bottom of the glass tube, so that the whole length of the wire was protected by oil.

In order that the stress acting on the wire may remain constant when the wire is stretched, the force applied to the wire must diminish so as to be at every stage proportional to the cross-section of the wire. Two methods of satisfying this condition were tried. The first method was that of the hyperbolic weight,* in which the stress is applied by a weight having the form of a hyperboloid of revolution. This sinks into a liquid as the wire extends, with consequent automatic reduction of the extending force. The friction at the pulleys, which have to be introduced to apply the force to the wire in an upward direction, was found to be troublesome, and a second method, in which friction was practically eliminated by the use of a knife edge instead of pulleys, was therefore devised for applying the stress, as follows.

An aluminium beam PH, fig. 1, length 40 cm., is supported by a knife edge B, and carries two plates F and C, one at each end. The plate C has a groove along its outer edge HK, the profile of HK being an arc of a circle with centre B. D is a thin steel wire resting in the groove and fixed to the adjusting screw E. The lower end of D is attached to the upper end of the wire to be stretched. F is the second plate, in the groove of which lies a thin steel wire supporting a weight W. The profile of the bottom of the groove PQR is made such that

* Andrade, 'Proc. Roy. Soc.,' A, vol. 84, p. 1 (1910).

the moment of the weight W about the axis through B is inversely proportional to the length of the wire undergoing stretch, which, on the assumption that

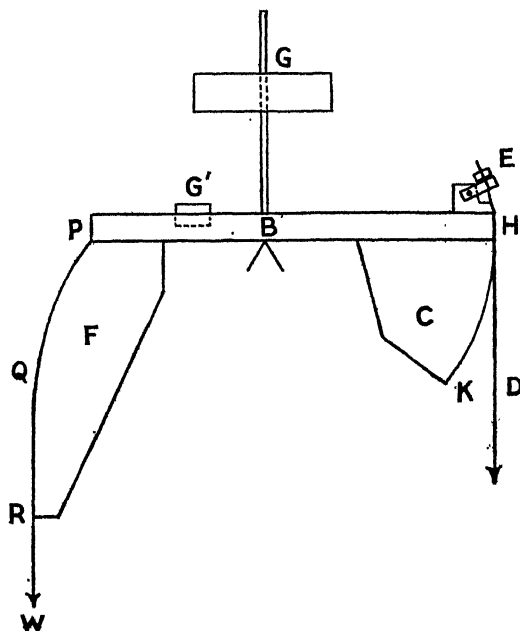


FIG. 1.

any change in density of the metal which may take place during stretch is negligible,* will make the stretching force proportional to the cross-section of the wire. Let p be the length BL of the perpendicular through B on to the vertical LQN through Q (fig. 2), and let this perpendicular make an angle θ

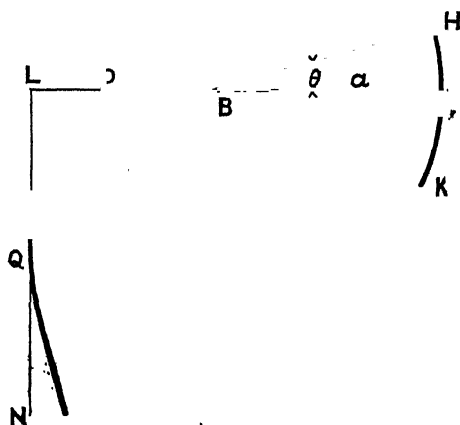


FIG. 2.

* See p. 358.

with a line, fixed in the beam, which is horizontal at the beginning of the experiment, when the wire is unstretched. Let a be the radius of the circular arc HK, centre B. Then if the initial length of the wire is l , its length at the moment considered is $l + a\theta$, and the required condition for the moment is

$$\frac{p}{a} = \frac{l}{l + a\theta},$$

which is a pedal equation to the required curve. The curve can easily be obtained graphically by drawing tangents, since the wire carrying the weight W is always vertical and tangential to PQR fig. 1.

In order to eliminate any other moments about B it was necessary to arrange the position of the centre of gravity of the whole arrangement to coincide with the knife edge B. This was done by means of two movable weights G, G' , fig. 1, which were adjusted and fixed in such positions that the time of swing about B was as long as possible.

An advantage of using this method of obtaining a constant stress is that the only specified quantity is the length l ; so that wire of any cross-section can be used with any stretching force consistent with the mechanical strength of the beam.

To prove the efficiency of the apparatus, and to ascertain how nearly the stress remained constant, the following test was made. A constant weight was supported by the wire D, fig. 1, and the angles at which the beam was in equilibrium when various larger weights were fixed to W were measured. From these data the ratio of the force acting on the wire with the beam at any angle to the force with the beam horizontal could be calculated. The angle of the beam enables the cross-section of the stretched wire to be calculated, and it was found that the stress was constant to within 1 per cent. up to an extension of 30 per cent. of the original length of the wire.

In the case of cadmium considerable difficulty was encountered in finding a method of fixing the ends of the wire in such a way as to give consistent mechanical and electrical contact. Soldering, which was the first method adopted, appeared to be unsatisfactory, as the resistance changed slowly even when no stress was applied. This effect, however, was afterwards found to be attributable to the water which surrounded the wire in the earlier experiments. As already stated, an oil bath was adopted after the preliminary experiments, but soldering was still found to be unsatisfactory, as the wire fractured near to the point of soldering under stresses that were too low to give satisfactory readings for the extension. It was discovered that a satisfactory method of

fixing the ends of the wire was to fuse them into spheres whose radius was two or three times that of the wire, and to grip these spheres in small four-jaw chucks. The ends were fused by heating in a small gas flame with a small quantity of soldering flux. Unless flux was used, the metal would not fuse satisfactorily, but retained its cylindrical shape, even when raised to a temperature well above its melting point. It was found after most of the readings had been taken that wires prepared in this way and subsequently soldered, gave quite satisfactory results.

With both tin and copper, no special precautions were necessary and ordinary soldering was employed.

3. Measurement of Length.

To obtain simultaneous values of the change in length and resistance of the wire a continuous record of the extension was made automatically; the resistance measurements were made at regular intervals which were marked on the record by the depression of a key. The recording of the extension was effected as follows. A light brass frame carrying a transparent scale was attached to the upper end of the wire whose extension was to be measured, the upward pull being applied by the fine steel wire D, fig. 1. The brass frame was constrained to move in a vertical plane by means of guiding grooves; the friction introduced by these was measured and found to be negligible. By means of a pointolite lamp P and lenses L_1 , L_2 , fig. 3, an image of the scale, T, by

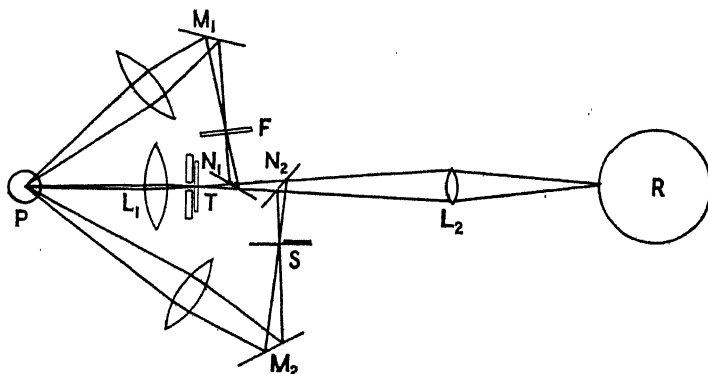


FIG. 3.

limited in breadth by an opaque vertical slit, was thrown on to a sheet of sensitive bromide paper wrapped tightly round a clock-work drum R. As the drum rotated the images of the scale divisions traced out a series of parallel lines, inclined to the horizontal more or less according as the wire was stretching

faster or slower. To facilitate the measurement of the extension an image of a fixed scale F was simultaneously recorded on the drum by means of a mirror M_1 and a half-silvered mirror N_1 , while for the time measurements an image of the slit S was similarly thrown on the drum by means of the mirror M_2 and the half-silvered mirror N_2 . This slit was obscured by a shutter which was opened once every minute by the agency of a clock actuating a mercury contact. The optical paths of the rays forming the images of the two scales and of the slit were carefully adjusted to equality.

Both scales were prepared photographically to show transparent markings on a black ground, each division of the fixed scale being approximately 0.25 mm., while the divisions of the moving scale were such that 24 of them were exactly equivalent to 25 of the fixed scale. The superposed images formed on the drum showed a magnification of about four times. This combination of scales enabled readings to be taken to $1/25$ th of a scale division, i.e., to 0.01 mm. by a coincidence method carried out as follows.

Part of a record developed on the bromide paper is represented in fig. 4. The lines AA' , BB' , CC' , ..., are the timing marks, representing images of the vertical slit S opened periodically by means of the clock contact. The horizontal lines KK' , LL' , MM' , ..., are the images of the divisions of the fixed scale F and the oblique lines are due to the moving scale T . The displacement of these lines perpendicular to KK' at any given instant is to be measured. To read the extension occurring between two timing marks, say, CC' and DD' , any one of the horizontal lines, say, MM' , is chosen as a standard line. The extension is measured as a whole number of units of 0.25 mm., represented by the number of intervals, counted along MM' , between consecutive oblique lines, a fractional part which is the extension subsequent to the last intersection and a similar fractional part at the beginning. Lines of coincidence mn and pq are drawn cutting the timing mark DD' , at which the extension is to be found, in g and h . It follows from the ratio of the scale units that two consecutive lines of coincidence, e.g., mn and pq , must cut a timing mark, e.g., DD' at a vertical separation of 25 of the divisions between consecutive lines MM' , NN' . The horizontal distance ad corresponds to an extension of 0.25 mm., and it is required to find the extension corresponding to ab .

Now

$$ab/ad = gb/gh,$$

but gh is 25 scale divisions, so that if gb is expressed as a number of scale divisions, then

$$ab = gb \times \frac{ad}{25}.$$

The distance ad corresponds to 0.25 mm. and hence gb , measured in terms of the unit separation of the horizontal lines, represents the extension in hundredths of a millimetre.

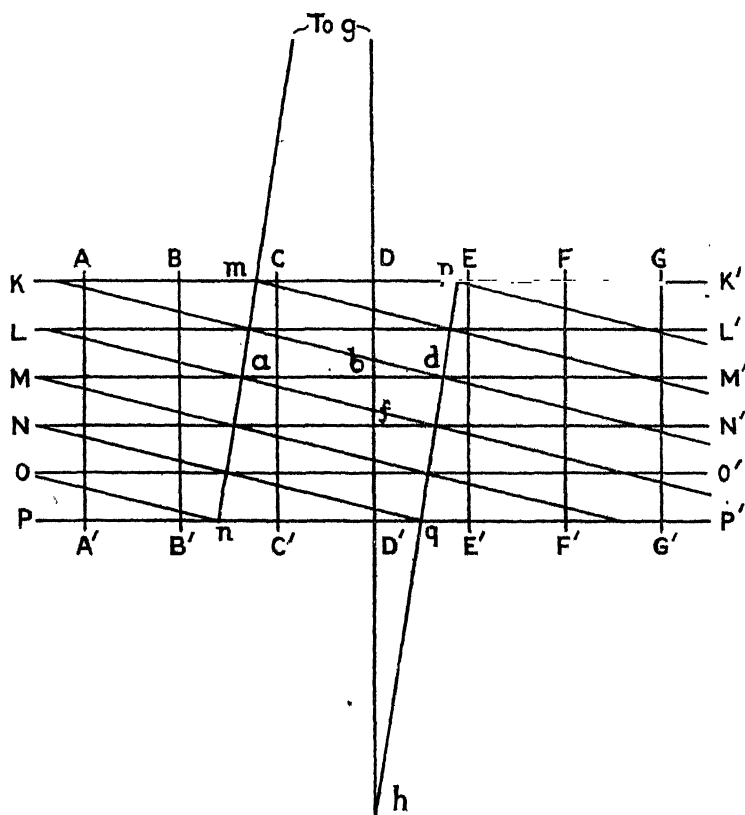


FIG. 4.

In this way the total extension at any timing mark can be found to within 0.01 mm. It is also possible to read off the slope of the oblique lines (*i.e.*, rate of stretch) directly, as the inclination of the line of coincidence to the horizontal is 24 times that of the lines given by the moving scale.

A correction had to be made for the fact that the spacing of the moving scale was not exactly 0.25 mm., but was actually 0.261 mm. A table was prepared to facilitate the reduction of readings.

4. Resistance Measurements.

The resistance of the wire was measured with a Callendar-Griffiths bridge of high accuracy used by Mr. N. Eumorfopoulos in his standard determinations

of the boiling point of sulphur, and kindly lent to us by him. The clamps holding the wire were used as leads, and their resistances, together with those of the leads to the bridge, were found by taking a reading on the bridge with a thick copper wire in the clamp in place of the cadmium or other wire. The resistance of the copper wire was only 0.006 of that of the cadmium wire, allowance for which being made the true resistance of the cadmium wire was at once found. By the use of compensating leads to eliminate the effect of temperature change and by lagging the bridge wire with cotton wool, resistances could be read consistently to 0.0002 ohms, which corresponds to a length of 0.1 mm. of the bridge wire.

The method of taking readings was as follows. The wire, previously annealed was fixed in place and the optical system adjusted. A weight, say, 50 gm., just sufficient to keep the wire taut, was then suspended at W and the clock-work drum carrying the sensitive paper was started. The initial length of the wire was then measured by a cathetometer to within 0.1 mm., and the resistance found by setting the bridge slider so that no change of deflection occurred on reversing the battery. When these two measurements had been made, a key, in parallel with the clock contact, was depressed for 3 seconds, producing a thick time line on the record to mark the instant at which the resistance was known.

The final weight (300 to 1000 gm.) was then applied, and the bridge slider set to a definite selected value. The battery was repeatedly reversed, with the object of eliminating thermo-currents, until no change of deflection occurred on reversal, at which instant the time key was depressed. This process was repeated at intervals throughout the extension. With the help of these time marks, the exact length corresponding to a known resistance could be read off from the photographic record, as explained.

The accuracy of the readings was as follows :—

- (a) The extension was measured to within 0.001 cm. ; the initial length of the wire was about 20 cm., and was itself measured within an accuracy of 1 in 2000. Thus the stretch, measured as a percentage of the original length, could be measured to 1 in 2000, while on a total stretch of, say, 20 per cent. of the original length, the stretch at any moment could be measured to within 1 in 4000 of the total stretch.
- (b) The limit of accuracy in the resistance measurements was about 1 in 5000, the total resistance being of the order of 0.1 ohm. With the rates of stretch obtained a change of resistance of the magnitude of the possible error took place in about 2 seconds. The estimated error in

recording the time at which the resistance was measured was about 3 seconds, so that the resistance corresponding to a given length registered on the record was probably obtained correctly to 1 in 3000.

EXPERIMENTAL RESULTS.

5. Variation of Length with Time under Constant Stress.

The variation of length with time of annealed cadmium wires under constant stress for temperatures ranging from 0° to 100° C., and for stresses varying from 180 kilogram wt./cm.² to 600 kilogram wt./cm.² gave curves of the general type to be expected on the previous work of Andrade (*loc. cit.*), on other metals. A typical example obtained at 66° C. with a load of 500 gm., *i.e.*, a stress of 301 kilogram wt./cm.², is given in the following table, where the observed extension is compared with that calculated from the formula

$$l = l_0 (1 + t^{1/3}) e^{kt}. \quad (1)$$

$$\beta = 0.0288. \quad k = 0.001895. \quad l_0 = 1.0053.$$

<i>t.</i>	Length (experimental).	Length (calculated).	E. — C.
0	1.000	1.000	0
1.20	1.0357	1.0384	-0.0027
2.17	1.0451	1.0471	-0.0020
4.75	1.0631	1.0635	-0.0004
8.48	1.0814	1.0816	-0.0002
12.72	1.0992	1.0991	+0.0001
19.98	1.1261	1.1258	+0.0003
25.20	1.1438	1.1435	+0.0003
30.45	1.1604	1.1609	-0.0005
35.60	1.1778	1.1774	+0.0004
43.33	1.2018	1.2018	0.0000
45.70	1.2092	1.2092	0.0000

The agreement is very good, probably within experimental error, except during the first 2 or 3 minutes, where the formula shows an extension a little greater than that observed. The discrepancy is not very great, as appears clearly from fig. 5, which exhibits the results embodied in the above table, and is

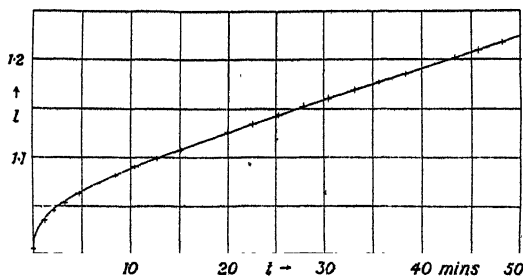


FIG. 5.

given to illustrate the general form of the curve. The continuous line represents the formula, the crosses being the observed points. While the extension at this early stage may possibly be slightly affected by irregularities in the first application of the stress, the weight being lowered into position by hand, other results seem to establish the fact that the formula does not represent the extension for cadmium quite exactly during the first minute or two. For high temperature and large stresses, in particular, while a good fit can be obtained it is necessary to take for this a value of l_0 which is less than the initial length of the wire. In this paper the formula is used to separate out that part of the flow which diminishes steadily with time, and there is no doubt that, apart from the first minute or two, it does represent the facts exceedingly closely. In place of $t^{1/3}$ is probably required a function, possibly an exceedingly complicated one depending on temperature and stress, which is distinctly less than $t^{1/3}$ for very small values of t , but which agrees closely with $t^{1/3}$ over a large range.

The total β -stretch in the above case is about 10 per cent. of the original length. Occasionally somewhat greater values were obtained. The maximum total extension was of the order of 30 per cent. of the original length.

6. *Specific Resistance.*

The specific resistance is, of course, obtained from the measured resistance r by dividing by the length l of the specimen, and multiplying by the cross-section s . If the density of the metal may be taken as constant throughout the flow, then the product ls is constant, and a quantity proportional to the specific resistance is obtained by dividing r by l^2 . It is known that the density of a single crystal is not the same as that of the polycrystalline metal, so that it is probable that the stretching of an annealed polycrystalline wire has some effect on the density. For the complete change from single crystal to polycrystal the effect is, however, of the order of 1 in 3000 for, *e.g.*, iron, nickel and aluminium,* so that it may fairly be estimated that the change in density during stretch is negligible compared to experimental error.

The expression $r/l^2 = R$ may therefore be taken to give the resistivity R , and on this basis values of $\Delta R/R_0$ were calculated corresponding to various values of l at different constant temperatures, ΔR being the change, and R_0 the initial value, of R . For cadmium there was, in all cases, a decrease of

* Seiji Kaya, 'Kinzoku no Kenkyu,' vol. 5, p. 10 (1928), quoted by H. C. H. Carpenter, 'Not. Proc. Roy. Inst.,' vol. 26, p. 267 (1930).

resistivity, so that ΔR is negative. Fig. 6 shows $-(\Delta R/R_0)$ plotted against l for various temperatures and loads, the initial cross-sectional area of the wire being 0.166 sq. mm. in all cases. The curves show clearly certain general features. During the immediate extension there is no change of resistivity, although in the case of the experiment at 0° this immediate extension amounted to 8 per cent. of the original length, and in the smallest case shown nearly 2 per cent. During the intermediate extension the resistivity *decreases* markedly, but during the last phase of the extension, when the flow is tending towards the constant rate expressed by k , the resistivity is tending to a constant value. It would appear, then, that to the three phases into which the flow

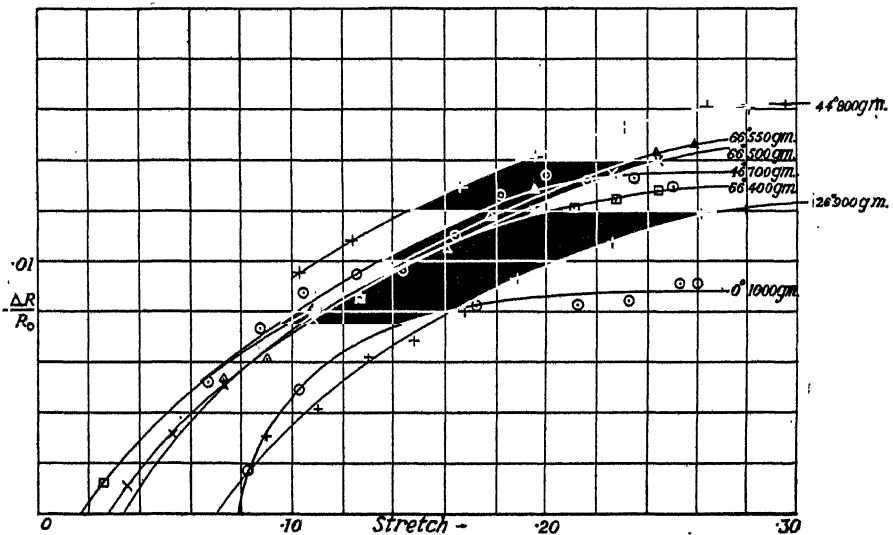


FIG. 6.

has been analysed correspond three distinct types of behaviour of the resistivity, but that, as appears from comparison of fig. 5 with the corresponding curve of fig. 6, the immediate extension is greater than is given by the formula, that is, by the $t^{1/3}$ law. This is in agreement with the argument already adduced. In view of the theoretical discussion to be given later, where the change of resistivity is connected with the β -flow, this point is of importance. The clear experimental evidence for constancy of resistivity when the wire has reached the state where it flows linearly with the time should also be noted.

The detailed course of the different curves of fig. 6 is somewhat complicated. For instance, the three curves for 66° cross one another. These points are not

discussed here since it will be argued later that the significance of the experimental results is better revealed by a different method of treatment.

The experiments were also carried out with copper and tin. In the case of copper, a load of 400 gm. applied to annealed copper wire, S.W.G. 36 (stress 1368 kilogram wt./cm.² at 100° C.) gave a stretch curve of the same form as that found for cadmium, the final extension being 7.5 per cent. of the initial length. The value found for the specific resistance was constant to within 0.1 per cent., the variations, of less than this amount, being quite irregular. Hence if any change of specific resistance does occur it is certainly less than one-tenth of the smallest change found for cadmium. The conclusion that there is no change of specific resistance in the case of copper agrees both with the results of other observers* and with other considerations cited in the discussion of results. The work with copper also goes to show that there is no systematic error in the readings, such as might be caused by the incorrect calibration of the bridge wire or of the moving scale.

An attempt was made to apply the same method to aluminium and tin but a difficulty was encountered in fixing the ends of the wire so as to make contacts of constant resistance. In the case of aluminium both clamping and soldering with a special aluminium solder were tried, but in all cases the resistance increased irregularly with time even when no tension was applied to the wire. This change of resistance was probably due to the formation of a film of oxide on the aluminium. In the case of tin a similar change of resistance, attributed to changes associated with the fixing of the ends, was detected. Since for this reason it proved impossible to measure the change of specific resistance during the extension, the final value of the specific resistance after stretching was compared to the initial value in order to determine whether any change had taken place.

The potential method was used, which was also employed for the measurement of specific resistance of single crystal wires (described in a paper to be published shortly). The wires whose resistances w_1 and w_2 are to be compared are arranged in series, being soldered to massive copper leads L, M, N, fig. 7. Each pair of potential contacts A, B and C, D consists of two safety razor blades of the three-hole type, fixed parallel and at about 8 cm. apart by glass rods passing through the holes and cemented. The contacts were held gently to the wires by rubber bands. Leads soldered to the blades pass to four mercury cups E, F, G, H, any one of which could be connected through a

* Cited by H. C. H. Carpenter in an unpublished lecture.

sensitive galvanometer to a sliding contact moving on a potentiometer wire ST. The potential drops p_1 and p_2 over AB and CD were measured on this wire in the usual way. The wires, whose resistances were being compared, were then cut off between the potential contacts by pressure on the razor blades and weighed, their masses being m_1 and m_2 . The ratio of the two resistances is then given by $w_1/w_2 = p_1/p_2$ and since

$$\sigma l = m,$$

where σ is the density

$$R_1/R_2 = p_1 m_1 l_2^2 / p_2 m_2 l_1^2,$$

where R_1 and R_2 are the specific resistances. The lengths l_1 and l_2 were very nearly the same. To eliminate the effect of any possible inaccuracy in the measurement of the length a second experiment was always carried out with stretched and unstretched wires interchanged, and the mean taken. This mean is quoted as the ratio.

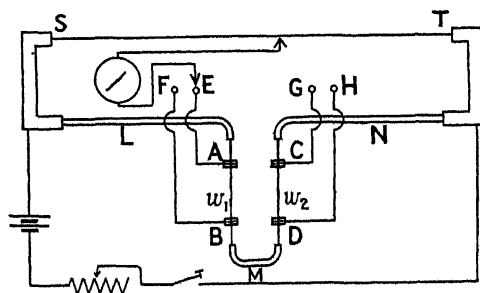


FIG. 7.

With a total extension of 5 per cent. the ratio stretched to unstretched for the resistivity was 0.999 for a particular stretched wire, and for a different stretched wire 0.998, giving a mean value of 0.9985. The estimated accuracy was ± 0.002 , so that, within experimental error, there is no change of resistivity for aluminium.

In the case of tin it was found that the change of specific resistance was in the opposite sense to that of cadmium, *i.e.*, the specific resistance *increased* during the second phase of the extension. As an example of the magnitude of the change, the case of an 18 S.W.G. wire at 100°C. may be cited, with initial load 1200 gm., which gives a stress of 1028 kilogram wt./cm.². The total extension during the second phase amounted to about 17 per cent. of the original length, during which the specific resistance increased by 2.3 per cent. During the first and during the third phase of the extension the

specific resistance remained unaltered as with cadmium. In a second experiment at the same temperature a total extension of 15 per cent. led to an increase of 2.4 per cent. in specific resistance.

DISCUSSION OF RESULTS.

7. *Resistance of Polycrystalline Wires in terms of Single Crystal Resistances.*

A general explanation of the results obtained with polycrystalline wires is to be sought in terms of the resistances of the crystallites of which they are made up. It is clear that this will be easier to find if, in the case of the soft metals in question, the crystallites alone need be considered, special effects at the interfaces being either non-existent or negligible. Information as to the magnitude of these possible effects can be obtained by considering the resistance of an unstrained polycrystalline wire in terms of its components.

For this purpose we suppose that the wire can be treated as an assemblage of small cubical blocks, each of which is a single crystal, the axes of crystal symmetry in the different blocks being oriented at random, and further that a layer of such crystallites, normal to the axis of the wire, can be treated as a number of resistances in parallel, since this will make the equipotential surface a plane normal to the wire. We further assume that there is no special resistance at crystal boundaries. The resistance of a single non-cubic crystal of a metal with properties completely symmetrical about a preferred axis of symmetry is given by the formula*

$$\begin{aligned} R &= R_{\parallel} \cos^2 \phi + R_{\perp} \sin^2 \phi \\ &= R_{\perp} + (R_{\parallel} - R_{\perp}) \cos^2 \phi, \end{aligned} \quad (2)$$

where ϕ is the angle which the axis of symmetry makes with the direction of the current and R_{\parallel} and R_{\perp} are the resistances along the unique axis of symmetry and normal to it respectively. The fraction of the crystallites for which the axis makes an angle between ϕ and $\phi + d\phi$ with the axis of the wire is, on the assumption of random distribution, $\sin \phi d\phi$, and therefore the resistance R_0 of the polycrystalline wire is given by

$$\begin{aligned} \frac{1}{R_0} &= \int_0^{\pi/2} \frac{\sin \phi d\phi}{R_{\perp} + (R_{\parallel} - R_{\perp}) \cos^2 \phi} \\ &= \frac{1}{\sqrt{R_{\perp}(R_{\parallel} - R_{\perp})}} \tan^{-1} \sqrt{\frac{R_{\parallel} - R_{\perp}}{R_{\perp}}}, \end{aligned}$$

* Voigt, "Lehrbuch der Krystallphysik" (1910). See Hume-Rothery, "The Metallic State," p. 7 (1931).

or

$$R_0 = \sqrt{R_{\perp}(R_{\parallel} - R_{\perp})} / \tan^{-1} \sqrt{\frac{R_{\parallel} - R_{\perp}}{R_{\perp}}},$$

and

$$\frac{R_0}{R_{\perp}} = p / \tan^{-1} p,$$

where

$$p^2 = \frac{R_{\parallel} - R_{\perp}}{R_{\perp}}.$$

Bridgman gives for cadmium the values $R_{\parallel} = 8.30$, and $R_{\perp} = 6.8$, while Grüneisen and Goens* give $R_{\parallel} = 7.79$ and $R_{\perp} = 6.54$. With Bridgman's values the above equation gives $R_0 = 7.27$, while with Grüneisen and Goen's values $R_0 = 6.94$. In the course of some later work, shortly to be published, we found $R_{\perp} = 993$ in terms of $R_0 = 1000$. Taking Bridgman's value of R_{\perp} this gives $R_0 = 7.24$ as against the calculated 7.27, while with Grüneisen and Goen's value of R_{\perp} we have $R_0 = 6.96$ as against the calculated 6.94. The formula therefore gives the resistance of the polycrystalline wire in terms of R_{\parallel} and R_{\perp} within less than 0.5 per cent., the discrepancy being within the limits of the experimental errors.

The hypothesis that the resistance can be considered as made up of the resistances of the elementary crystallites, without special interfacial resistance, may therefore be considered reasonable. There is, of course, the possibility that the polycrystalline wires, while containing crystallites set at all angles, have, owing to the method of preparation, one or more preferred directions in which more crystallites are oriented than would be given by a purely random distribution. This is a point that can only be settled by X-ray examination, and Dr. R. E. Gibbs now has this and other points under examination.

8. Quantitative Explanation of Change of Resistance during Extension.

As a preliminary to the consideration of the change of resistance during extension in terms of the behaviour of the crystallites, certain properties of metal single crystals may be recalled. When a single crystal wire is extended, glide takes place along a certain set of parallel crystal planes; if, as with certain classes of crystals, this set is one of a group of planes which are crystallographically equivalent, then, of course, slipping takes place along that set of the group which is, mechanically, the most favourably disposed. In addition, glide takes place in a certain direction in this plane, or in one of a certain group of equivalent directions. Thus, to take an example, with cadmium, which crystallises

* 'Z. Physik,' vol. 26, p. 250 (1924).

in the hexagonal form, the glide takes place in the basal, or (0001) plane, and in the direction of a digonal axis from the centre to one of the vertices of the hexagon. The direction actually chosen by the glide will be the one among the preferred axes which makes the smallest angle with the direction of the applied force; since the digonal axes make an angle of 60° with one another the direction of glide cannot be more than 30° from the line in the glide plane which makes a minimum angle with the direction of the force.*

Thus consider a single crystal wire of cadmium, in which the axis of crystal symmetry makes initially an angle ϕ_0 with the axis of the wire, while the glide planes make an angle θ_0 with the axis of the wire; here θ_0 and ϕ_0 are complementary. When the wire is stretched by a force applied in the direction of the axis the planes glide parallel to themselves, but the axis of the wire in which glide has taken place is constrained by the force to remain in the original direction of the axis. The result is that ϕ_0 and θ_0 take new values ϕ and θ which are still complementary. It can easily be shown† that

$$l/l_0 = \frac{\sin \theta_0}{\sin \theta} = \frac{\cos \phi_0}{\cos \phi}, \quad (3)$$

where l , l_0 are the final and original lengths of the wire. Extension by glide in the single crystal of cadmium therefore leads to a rotation of the unique set of glide planes.

The results of which an explanation is sought are that there is no change of specific resistance with copper or aluminium at any stage of the extension, nor during the immediate extension or the final flow at constant rate with any of the metals. With cadmium and tin, however, there is a definite change of specific resistance when wires are extended at atmospheric or higher temperatures, the change (which is a decrease for cadmium but an increase for tin) being closely associated with the β -flow.

A polycrystalline wire can be considered as an assemblage of crystallites in which the axes of crystal symmetry are oriented at random. When the wire is stretched three possible mechanisms can be imagined: (a) that the crystallites break up into smaller fragments, which move without rotation so as to

* See, e.g., Taylor and Elam, 'Proc. Roy. Soc.,' A, vol. 108, p. 28 (1925); Mark, Polanyi and Schmid, 'Z. Physik,' vol. 12, p. 58 (1922); Georgieff and Schmid, 'Z. Physik,' vol. 36, p. 759 (1926); and Boas and Schmid, 'Z. Physik,' vol. 54, p. 16 (1929). For collected accounts see Wien-Harms, *Handbuch der Experimentalphysik*, vol. 5 (Plastische Verformung, by G. Sachs) (1930), and Masing and Polanyi, 'Ergebn. exakt. Naturwiss.,' vol. 2, p. 177 (1923).

† E.g., Ewald, Pöschl and Prandtl, "The Physics of Solids and Fluids," p. 120 (1930).

accommodate their positions to the bulk deformation of the wire. The necessary accommodations may lead to slight deformations, of the cylindrical form, and, in fact, the surface of a stretched polycrystalline wire generally appears somewhat rough; (b) that twinning takes place; (c) that the individual crystallites become extended in the direction of the axis of the wire, not necessarily all by the same amount. In this case, if the glide takes place entirely upon a unique set of glide planes, as it undoubtedly does in single crystals of cadmium, or even if it is mainly upon such a set of glide planes, there must be a change in the angle ϕ , determined by the extension, that is, a rotation of the crystalline axis of the crystallite.

In the first case no change of resistance is to be expected, as far as the crystallites are concerned, and it has already been shown that with cadmium, at any rate, the resistance can be wholly accounted for by the crystallites, without interfacial considerations.

Consideration of the second case will be postponed, since examination of polycrystalline wires, stretched at atmospheric or higher temperature, has failed to reveal twinning. The examination was carried out by polishing the surface of the wire, and etching with dilute nitric acid, which shows up the crystal boundaries, but not slip planes. That this process will also reveal twinning if it has taken place is evidenced by cadmium wires stretched at low temperatures of -180° C. which show extensive twinning, as recorded in the last section of this paper.

In the third case we must expect a change of resistance for all metals in which there is a unique axis of crystal symmetry, for in such metals the electrical resistance depends upon the inclination of the axis to the direction of flow of the current. The results of the experiments here described lead to the conclusion that, during the β -flow, the extension takes place mainly, if not wholly, by extension of the crystallites, accompanied by rotation of the axes. It may be mentioned that an early study of the mechanical aspect of the flow in soft metals led to the conclusion that the β -flow was due to crystalline rotation.*

A strong quantitative confirmation of this view is furnished by the comparative behaviour of cadmium, tin, copper and aluminium. In the hexagonal crystals of cadmium the glide planes are normal to the unique axis of symmetry; in the tetragonal crystals of tin they are parallel to the unique axis.† In the case of both metals the resistance in the direction parallel to the unique axis

* Andrade, 'Proc. Roy. Soc.,' A, vol. 84, p. 1 (1910).

† Polanyi, 'Naturwiss.,' vol. 16, p. 209 (1928).

is greater than in the normal direction.* The effect of stretch of a crystal or crystallite of either metal is to tend to set the glide planes parallel to the direction of stretch, which means, in the case of cadmium, setting the unique axis normal to the direction of stretch, in the case of tin, bringing the unique axis parallel to the direction of stretch. Hence in the case of cadmium there would be a decrease of specific resistance accompanying β -stretch, in the case of tin an increase, which is what is observed. In the case of a cubic crystal, like copper, the resistance is independent of the direction, so that there should be no change of resistance.

There is, then, good experimental evidence for believing that, in the case of soft metals, at any rate, the first immediate stretch takes place by break-up of the crystallites and movement of the fragments without distortion, by slipping at the boundaries; that in the slow stretch which follows, rotation of the crystallites takes place, with consequent change of specific resistance, and that in the final stage of steady flow the rotation is complete, and that slipping without rotation occurs. It may be objected that there is a difficulty, of a geometrical nature, in imagining exactly how the adaptation at the boundary between individual crystallites takes place so that the metal may remain continuous. At present so little is known of the behaviour of a mass of plastic crystallites that a precise handling of the problem is not in question, but attention may be directed to certain arguments that seem to justify the assumption of rotation. In the first place, it is well known that in the case of wire drawing of many metals, *e.g.*, copper and aluminium, a preferential orientation of certain crystal planes takes place, although no experimental results on wire drawing with cadmium are known to us. Mark has shown by the X-ray method a strong preferential orientation in a drawn tin wire.† Rolling is also known to produce a regular texture and, for instance, in the case of zinc and cadmium about 70 per cent. of the crystals lie, after rolling, with their hexagonal axes making an angle between 50° and 70° with the normal to the sheet.‡ Although the deformation in our case is not nearly as severe as that produced by rolling or wire drawing, work on these methods shows clearly that mechanical deformation, can lead to orientation with the necessary adaptations.

Again, in soft metals of the kind in question the crystal planes can bend so as to adapt themselves to geometrical constraints, as evidenced, for example, by a cadmium single crystal wire gripped at one end, and extended by a load.

* See, *e.g.*, Bridgman, *loc. cit.*

† Wien-Harms, "Handbuch der Physik," vol. 7, p. 311 (1928).

‡ Mark, 'Z. Kristallog.', vol. 61, p. 75 (1925).

In the part where the wire is gripped the planes retain their original inclination; at some little distance from this part the inclination of the planes has a different value, and in the intermediate region the adaptation between the two inclinations is effected by a bending of the crystal planes. Again, the picture put forward does not exclude the possibility that the general movement of the undistorted crystallites may be accompanied by a slipping and rotation in a relatively small fraction of the total number, nor that general slipping and rotation may be accompanied by a breaking-up of a few crystallites. The variation of the specific resistance with extension cannot be calculated without special hypotheses as to the way in which glide is initiated, continues and ceases in the individual crystallites of the polycrystalline wire. While on the basis of such special hypotheses we have obtained fairly close agreement with experiment, it is well, perhaps, to reserve such calculations until the hypotheses have further experimental support, which is now being sought.

Meanwhile the experimental results can be expressed in terms of a mean angle $\bar{\phi}$, chosen so that, if in all the crystallites the angle made by the crystal axis with the axis of the wire had the same value $\bar{\phi}$, then the resistance would be the same as it is in the actual polycrystalline wire. Any preferential orientation in the polycrystalline wire modifies the value of $\bar{\phi}$ so defined.

By definition,

$$R = R_{\perp} + (R_{\parallel} - R_{\perp}) \cos^2 \bar{\phi} = R_{\perp} (1 + p^2 \cos^2 \bar{\phi}). \quad (4)$$

For an unstretched polycrystalline wire, with random orientation,

$$\frac{R_0}{R_{\perp}} = p / \tan^{-1} p = 1 + p^2 \cos^2 \bar{\phi}_0$$

or

$$\begin{aligned} \cos \bar{\phi}_0 &= \frac{1}{p} \sqrt{p / \tan^{-1} p - 1} \\ &= 0.5617 \text{ with Bridgman's value of } p. \\ \bar{\phi}_0 &= 55^{\circ} 50'. \end{aligned}$$

For any given stretched polycrystalline wire $\bar{\phi}$ is obtained from the experimental value by substituting the value of R in equation (4).

9. Analysis of Experimental Results.

There is, it appears, good reason to associate the change of resistance in metals of non-cubic structure with the β -flow. By fitting to the length-time curves the equation $l = l_0 (1 + \beta t^{1/3}) e^{kt}$ the appropriate value of β can be found

for each (constant) stress and temperature, and so for each observed value of the specific resistance, taken at a time t during the flow, the corresponding value $\beta t^{1/2}$ of the β -stretch can be found. In this way figures were obtained from which the curves of figs. 8 and 9, exhibiting $-\Delta R/R_0$ for cadmium as

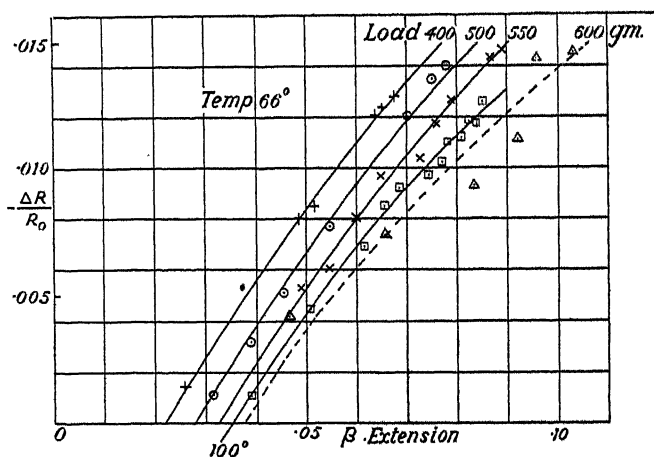


FIG. 8.

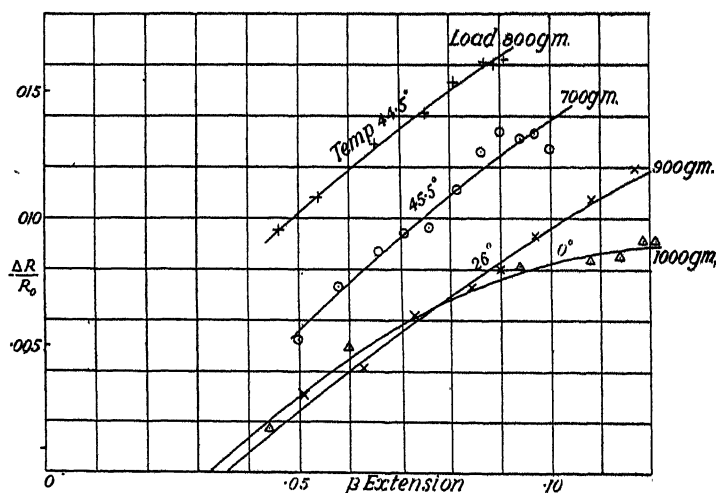


FIG. 9.

ordinate against β -stretch as abscissa were plotted. Each curve refers to a wire extended at fixed temperature and constant stress. The curves of fig. 8 were all obtained at temperature 66°C ., with the exception of the curve marked 100° , obtained at that temperature, the initial load being indicated on each curve. The initial cross-sectional area of the wire being 0.166 sq. mm. the

constant stresses corresponding to 400, 500, 550 and 600 gm. weight are 2.41, 3.01, 3.31, 3.61×10^5 gm. wt./sq. cm. The points obtained with the 600 gm. load are very irregular; the flow in this case was the most rapid observed in the course of this investigation, the wire stretching from an initial length of 20 cm. to a final length of about 25 cm. in 13 minutes. On account of its uncertainty the curve is indicated by a broken line. It is difficult to fix β closely for the rapid flow at 100° , and therefore the slope (but not the form) of this curve is not so certain as that of the other curves. Two of the curves of fig. 9 were obtained at about 45° , one at 26° and one at 0° C.

The nearly straight nature of the curves, and the parallel course of curves obtained for different stresses at the same temperature, is in striking contrast to the unsystematic character of the group of curves of fig. 6, where the same ordinate is plotted against the total stretch, and at once suggests that there is a real physical significance in the analysis of the flow into immediate stretch, β -flow, and final flow at constant rate, or viscous flow, and that it is, in fact, the β -flow which is closely bound up with the change of resistance. The curves of figs. 8 and 9 do not, in general, pass through the origin, that is, when the flow is analysed into immediate stretch and β -flow by the help of the $t^{1/3}$ formula there is, included in the β -flow, an immediate extension which produces no change of resistance. This may be taken to indicate that the extension during the first minute or so, expressed as the early part of $\beta t^{1/3}$, is not due to glide and rotation, but is of the same nature as the immediate stretch, for the first resistance measurements shown on, *e.g.*, the 400 and 500 gm. curves of fig. 8 were taken within the first minute or so. This suggests that the empirical $t^{1/3}$ law, while expressing accurately a physically distinct process, which we have indicated to be extension due to glide and rotation, for moderate and large values of t , does not closely represent the process during the minute or two following the first application of stress. This is the conclusion already reached in Section 5, in the discussion of the fitting of the length-time curves by formula (1), where it is pointed out that, in some cases, to fit the experimental results l_0 has to be taken less than the initial value of l , which shows that $\beta t^{1/3}$ is initially too large to represent the physical process in the first minute or two.

Accordingly, we proceed on the assumption that the immediate stretch, indicated in figs. 8 and 9, which takes place without change of resistance is due to a displacement of the crystallites without glide or rotation, and proceed to consider the rate at which resistance decreases with increase of β -extension, supposed due to glide and rotation. It is clear that the curves obtained at one temperature, *i.e.*, the 400, 500 and 600 gm. curves at 66° , or the 800 and 700 gm.

curves obtained at 45° , differ only in their intercept on the horizontal axis, so that a given extension due to glide and rotation is, at a given temperature, connected with a given change of specific resistance. From the change of specific resistance we can calculate $\bar{\phi}$. In this way the curves of fig. 10 have

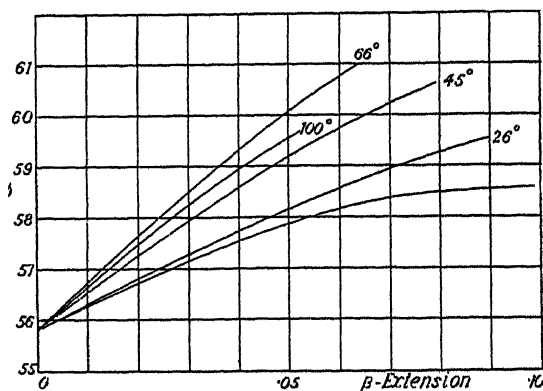


FIG. 10.

been obtained, which show $\bar{\phi}$ as a function of the extension, measured from the intercept on the horizontal axis in figs. 8 and 9, which, on our assumption, is due to rotation.

The curve for 100° represents only one experiment; it has been already pointed out that, on account of the very rapid flow, it is not possible to fix the β -extension accurately in this case. No great significance can be attached to the fact that the slope of this curve is less than that of the 66° curve. Leaving it out of account we see that, as the temperature falls, the change of $\bar{\phi}$ for a given extension diminishes, although for the 26° which separates the lower two runs the change is less marked than for the 20° intervals which separate the other curves. This can be simply interpreted if it is remembered that for crystallites for which ϕ_0 is large, *i.e.*, whose glide planes make small angles with the wire axis, a small change of angle corresponds to a large increase in length, while for those for which ϕ_0 is small a large change in angle is needed to give a small change in length. If for crystallites of small ϕ_0 glide takes place more readily at high temperatures than it does at low our results are explained.

If we can apply results obtained with single crystals to the crystallites in the wire, then glide takes place most easily in those for which the glide planes, and hence the crystal axes, make an angle of 45° with the wire axis. If we assume that glide takes place only in crystallites, for which the angle is in the neighbourhood of 45° , we can obtain an estimate of the change in resistance to be

expected in association with a given change in length. No doubt this is only a rough approximation, except in the early stages, but it affords a check as to whether changes of resistance of the magnitude observed by us are reasonably explained on the hypothesis of glide. Various more complicated assumptions give, as a matter of fact, not very different numerical results.

Suppose then, that on extension the axes of a fraction f of the crystallites, which initially made angles in the neighbourhood of 45° with the wire axis, change their directions by $d\phi$. We then have

$$\frac{\Delta l}{l} = f \cdot d\phi,$$

from

$$\frac{l}{l_0} = \frac{\cos \phi_0}{\cos \phi},$$

while, as an approximation,

$$\frac{\Delta R}{R_0} = \frac{p^2 f \cdot d\phi}{p/\tan^{-1} p} = p \tan^{-1} p \frac{\Delta l}{l}.$$

The numerical values being $p = 0.489$, $\tan^{-1} p = 0.439$

$$\frac{\Delta R}{R_0} = 0.206 \frac{\Delta l}{l}.$$

If

$$\frac{\Delta l}{l} = 0.01 \quad \frac{\Delta R}{R} = 0.00206 \quad \text{and} \quad \bar{\phi} = 56^\circ 25',$$

while if

$$\frac{\Delta l}{l} = 0.05 \quad \frac{\Delta R}{R} = 0.0103 \quad \text{and} \quad \bar{\phi} = 58^\circ 57'.$$

These are very slightly less than the experimentally found values for 45° C., as can be seen by consulting the curves of fig. 10. The changes of resistance are therefore of the right magnitude. A glide in the crystals for which the slip planes make smaller angles than 45° will tend to make the change of resistance less, in those for which the slip planes make larger angles will tend to make the change of resistance greater, for a given change in length.

10. Results at Low Temperatures.

The extension by glide and rotation, for which evidence is offered in this paper, is associated with the slow β -flow, and so apparently requires time for the necessary adjustments to take place. It will, therefore, characterise soft metals, or, speaking more correctly, metals at a temperature such that they are soft, and it is not contended that it will be found to accompany the extension

of harder metals. As the distinction drawn between the mechanism of flow in hard and soft metals has been criticised it seemed to us advisable to seek further experimental evidence on this point. If the contention is correct, then a cadmium wire at very low temperature, where the β -flow cannot take place, should behave differently, in respect of its changes of resistivity, from one at a temperature in the neighbourhood of atmospheric. Early experiments in which we attempted to produce extension at liquid air temperature failed because of difficulties found in gripping the end of the wire. Since the main work was completed it has, however, been found possible to hold the ends of the wire adequately, and a cadmium wire has been extended by 3 per cent. and more at both -180° C. and -78° C.

The results of these low temperature experiments were that while at -78° an extension of 3 per cent. produced a decrease of resistivity amounting to 0.7 per cent. (mean of two experiments), at -180° a stretch of 3 per cent. produced an *increase* of resistivity of 1.3 per cent., which seems at first astonishing. Examination of the wires after etching showed, however, that in the case of the wire stretched at -180° very extensive twinning had taken place. The bands of twinning in most of the crystallites made angles with the axis of the wire which were not far from 45° ; no bands either parallel to or perpendicular to the axis were observed. The increase of specific resistance can be explained in terms of these observations.

In the first place, working with single crystal wires of cadmium we have found that after primary stretch, which brings the hexagonal axis to a position nearly perpendicular to the force, is completed, a secondary stretch takes place if the stress is increased. This secondary extension is accompanied by twinning, and by a rise of resistivity. Cadmium twins on a (0112) plane, when the shearing stress in the plane exceeds a certain limiting value. The twinning plane makes an angle of not far from 45° with the original basal plane, so that after twinning the basal plane of half the twin stands nearly at right angles with the original basal plane, that is, nearly perpendicular to the axes of the single crystal wire. As the resistance parallel to the hexagonal axis is the larger one it is clear that twinning in such a stretched single crystal wire should increase the resistivity, as observed.

Turning now to the polycrystalline wire, where the directions of the axes are distributed at random, it might at first be supposed that twinning, which brings the axes nearly normal to their original positions, should not affect the resistivity. If, however, the twinning planes are mainly in the neighbourhood of 45° to the wire axis, as would be expected, and as is confirmed by

examinations of the twinned wire, then the crystals affected are mainly those in which the axis is either nearly parallel to or nearly normal to the wire axis. The latter class is clearly the more numerous in the case of random orientation. Since twinning in the former class leads to decrease of resistance, in the latter class to increase, it is clear that, on the whole, the effect of twinning should be to increase the resistivity, as is, in fact, observed.

The mechanism of yield is, then, essentially different for the tough wire at -180° and for the softer wire at atmospheric temperatures. The wire at -80° represents an intermediate case. The stretch at low temperatures has not been studied in detail, but it may be mentioned that a particular wire, extended at -180° by a very large load, under which it broke, showed a much greater increase of resistivity for unit extension than that quoted above.

We are much indebted to Dr. R. E. Gibbs for helpful discussion during the work. One of us (E. N. da C. A.) enjoys a grant for apparatus from the Imperial Chemical Industries Company, expenditure from which has been made in connection with this research, while the other (B. C.) has been in receipt of a grant from the Department of Scientific and Industrial Research. To both these bodies our best thanks are due.

11. *Summary of Results.*

(1) The specific resistance of certain typical metals has been determined at various stages of the plastic flow under large stresses.

(2) The specific resistance of metals which crystallise in the cubic system is unaffected by the flow.

(3) The specific resistance of metals which crystallise with a unique axis of symmetry does not change during two of the three stages into which the flow can be analysed, namely, during the immediate extension, and the final flow at constant rate. During the intermediate stage of flow at diminishing rate, called the β -flow, it changes by an amount of the order of 2 per cent. in extreme cases.

(4) The facts can be explained on the assumption that the crystallites slip, with consequent rotation of the unique axis, during the β -flow. In particular it follows at once on this hypothesis that with metals whose crystals have the slip planes parallel to the unique axis an increase of specific resistance with extension should be expected, while with crystallites which have the slip planes normal to the unique axis there should be a decrease, supposing that, as in the cases treated in this paper, the resistance is greatest along the unique

axis. This is what is found experimentally with tin and cadmium, which respectively represent the two classes of crystallite.

(5) The magnitude of the change of resistance during β -flow is in good accord with this hypothesis that β -flow is due to rotation of crystallites.

(6) Further evidence for the views put forward is afforded by experiments at low temperature, where there is marked immediate stretch, but no β -stretch. It has been found that, under these conditions, an increase of resistivity of cadmium is obtained as against the decrease at ordinary temperatures. This is caused by the extensive twinning which accompanies the stretch under these conditions.

(7) The results suggest that in all cases where flow is observed in solids the mechanism of the immediate stretch is essentially different from that of the β -flow, which is connected with a rotation of crystallites to a final position.

The Photochemistry of Phosphine.

By H. W. MELVILLE.

(Communicated by J. Kendall, F.R.S.—Received June 8, 1932.)

Within recent years the spectroscopy and the photochemistry of the hydrides of the non-metallic elements has been the subject of much study and the results explained with considerable success on the basis of the newer theories of photochemical reactions. In spite of the fact that the photochemistry of ammonia has attracted some attention, yet the investigation of the equally simple molecule of phosphine seems to have been entirely neglected. It is with the object of filling this gap in photochemistry that the present experiments have been undertaken.

The investigation naturally falls into three parts : (a) the absorption spectrum of phosphine, (b) the direct photochemical decomposition, (c) the photosensitised decomposition. Experimental data on the photo-oxidation are also of importance in view of the fact that the thermal oxidation of phosphine is a chain reaction.

In the present communication attention will be devoted to the photosensitised reactions, while the absorption spectrum and direct photo-reactions will, it is hoped, be dealt with in another paper.

The Photosensitised Decomposition.

Phosphine does not exhibit any absorption of visible or ultra-violet light until a wave-length of $230\text{ }\mu\mu$ is reached,* therefore, it is at once evident that, in principle, the mercury photosensitised reaction may be conveniently studied. The general plan of the investigation included experiments corresponding to those which had already been carried out with ammonia. In addition, it became necessary to extend this scheme as certain outstanding points required elucidation and extension.

Apparatus.

The arrangement of the apparatus employed is shown in fig. 1. The reaction tube was a long spiral of silica tubing 5 mm. in bore surrounding a simple form of mercury arc lamp. The lamp consisted of a slightly inclined silica tube 15 cm. long and 1 cm. in diameter fitted with side tubes and mercury reservoirs as is shown in the figure. The lamp and spiral were fitted into a brass tank through which water rapidly flowed for cooling purposes. The

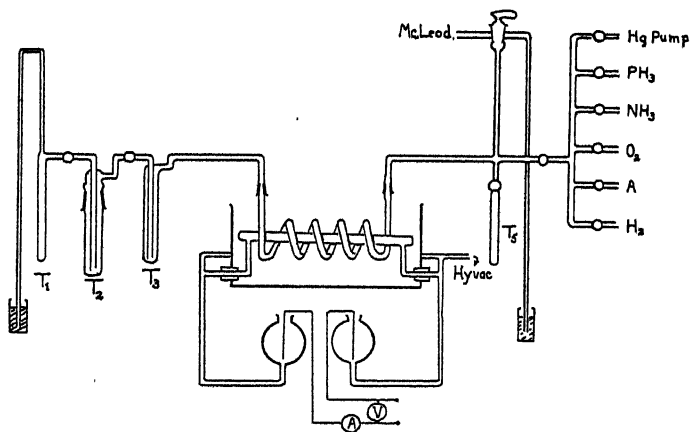


FIG. 1.

spiral was connected by means of silica-glass ground joints on the left side to a phosphine storage trap T₁, a trap T₂ for saturating the gas with phosphorus vapour, and T₃ which contained mercury in order to maintain the concentration of the vapour at a steady value in the spiral. T₃ was immersed in water contained in a vacuum flask. On the right side the spiral was attached (a) to storage bulbs and pipettes containing phosphine, ammonia, hydrogen, oxygen and argon; (b) to a mercury manometer, McLeod gauge and a mercury vapour

* Melville, 'Nature,' vol. 129, p. 546 (1932).

pump. Suitable liquid air traps and small reservoirs were connected to this part of the apparatus in order to facilitate the handling of the various gases used in these experiments.

During the preliminary work the brass box was filled with a 2.5 per cent. aqueous solution of potassium bromide in order to make sure that all light of $\lambda < 240 \mu\mu^*$ was filtered out from the radiation of the mercury lamp. In this case the solution was cooled by a coil of lead pipe in which a current of water circulated. After several hours' exposure the KBr was markedly decomposed. Later experiments showed that with unfiltered light from a water cooled arc the photosensitised was at least 100 times faster than the direct decomposition.

When it was found that the rate of the reaction depended to some extent on the surface-volume ratio of the reaction tube the arrangement of lamp and reaction tube was changed. A vertical water cooled lamp and interchangeable cylindrical insolation vessels were substituted for the horizontal lamp.

Since the experimental procedure varies with the type of experiment, details of the methods employed are most conveniently given with the numerical results.

Materials.

The phosphine was prepared from phosphonium iodide and potassium hydroxide, passed over soda lime, calcium chloride and phosphorus pentoxide collected in a liquid air trap after which it was fractionally distilled, the middle portion of the distillate being used. Ammonia was obtained by heating ammonium chloride and soda lime, dried with soda lime and fractionally distilled with solid carbon dioxide.

Products of Decomposition.

The first point to be settled about the photosensitised reaction is the nature of the end products. With the experimental arrangement described above using pressures of phosphine from 0.05 mm. up to 760 mm. an exposure of a few minutes or even seconds (at high phosphine pressures) sufficient to yield an amount of gas, non-condensable in liquid air, giving a pressure of about 0.1 mm. This gas could only be hydrogen. After a large number of exposures it was observed that the rate of decomposition gradually fell off and simultaneously a brown deposit was formed on the walls of the reaction tube. This

* 'Z. Elektrochem.,' vol. 35, p. 702 (1929).

brown deposit was most probably red phosphorus or a phosphorus hydride since it was easily removed by heating or by washing with bromine water.

If phosphine is decomposed according to the equation $4\text{PH}_3 = \text{P}_4 + 6\text{H}_2$, then the pressure of hydrogen produced should be 1.5 times that of the phosphine decomposed. Phosphine was introduced into the apparatus at a pressure which could be conveniently measured by the McLeod gauge, it was illuminated for a suitable length of time (in this case about 10 minutes, with 30 volts and 4 amps. in the arc) and the pressure again measured; the phosphine was then condensed out with liquid air in order to measure the pressure of the hydrogen. From these pressure readings, after making a small correction due to the fact that phosphine has a small vapour pressure at liquid air temperatures, it was then possible to calculate the percentage hydrogen produced. Table I gives a set of typical results.

Table I.

Initial pressure of PH_3 (mm.).	Total pressure after illumination.	Pressure of H_2 (mm.).	Percentage H_2 produced.
0.0550	0.0605	0.0200	92
0.0985	0.1055	0.0230	96
0.0645	0.0720	0.0245	95
0.1710	0.1925	0.0745	94
0.1015	0.1125	0.0390	93
0.0520	0.0585	0.0225	98
		Average	95

The average of 95 per cent. of the theory thus pointed to the conclusion that nearly all the phosphine decomposed formed phosphorus and hydrogen. The small deficiency in the amount of hydrogen is due, as will be shown later, to the clean up of part of the hydrogen on the walls of the insolation tube. Whether phosphorus vapour is first produced in the decomposition is not known, since under the influence of light from the mercury lamp the vapour is transformed into red phosphorus.

In spite of the fact that red phosphorus appears as the final product of the reaction, it was important to find if the addition of phosphorus vapour caused any decrease in the rate owing to the absorption of the $253.6\text{ }\mu\mu$ line of the mercury spectrum. A piece of yellow phosphorus was placed in T_2 which was kept at -80°C . while phosphine was passed into the spiral at about 2 mm. pressure. An exposure was made, the phosphine condensed out and the

pressure of hydrogen determined. After pumping out and refilling the spiral with the same pressure of phosphine, T_2 was warmed up to 15° (vapour pressure of yellow phosphorus 0.026 mm.) and a trap on the right side of the spiral cooled to -80° so that a constant stream of phosphorus vapour passed through the spiral. Experiments were made alternately with and without phosphorus vapour when it was found that only a slight diminution in rate occurred, *e.g.*, from 0.0250 mm./min. to 0.0245, 0.0295 to 0.0285, 0.0300 to 0.0295. The small influence of the phosphorus is, no doubt, due to the high absorption coefficient of the mercury vapour for the resonance line compared with that of phosphorus vapour; for example,* with a mercury vapour pressure of 0.001 mm. 95 per cent. of the resonance line is absorbed in a layer of vapour 2 cm. thick.

Effect of Hydrogen.

It is well known that molecular hydrogen is a particularly efficient deactivator of excited (2^3P_1) mercury atoms,† so that it would be expected that, as hydrogen accumulates in the phosphine during decomposition, the reaction velocity will fall off even if the pressure of phosphine is maintained at a constant value. This is clearly illustrated in fig. 2, and it is to be noted that the inhibition by hydrogen is more marked with small phosphine pressures. In order to discuss the matter quantitatively, it will be convenient at this point to propose a simple mechanism for the reaction and to find how far this mechanism can account for the experimental data. The rate of decomposition is given by

$$R = -\frac{d[PH_3]}{dt} = k_1[Hg'] [PH_3], \quad (1)$$

where k_1 is a constant, Hg' denotes the excited mercury atom in the 3P_1 state. Since the system is in a stationary state, then the following relationship exists

$$\frac{d[Hg']}{dt} = K - k_1[Hg'] [PH_3] - k_2[Hg'] [H_2] - k_3[Hg'] = 0, \quad (2)$$

or

$$[Hg'] = \frac{K}{k_1[PH_3] + k_2[H_2] + k_3}, \quad (3)$$

where K is proportional to the intensity of the illumination and the concentration of the mercury vapour. The third term of (2) takes into account

* Gaviola, 'Phil. Mag.', vol. 6, p. 1154 (1928).

† Stuart, 'Z. Physik,' vol. 32, p. 262 (1925); Zemansky, 'Phys. Rev.,' vol. 30, p. 919 (1930).

deactivation of Hg' by H_2 , while k_3 is the rate at which Hg' reverts to the 1^1S_0 state by radiation or by any other process which is independent of PH_3 and H_2 , *e.g.*, wall deactivation. Deactivation by radiation is probably the most important factor in this term owing to the short life (10^{-7} seconds) of the excited atom.

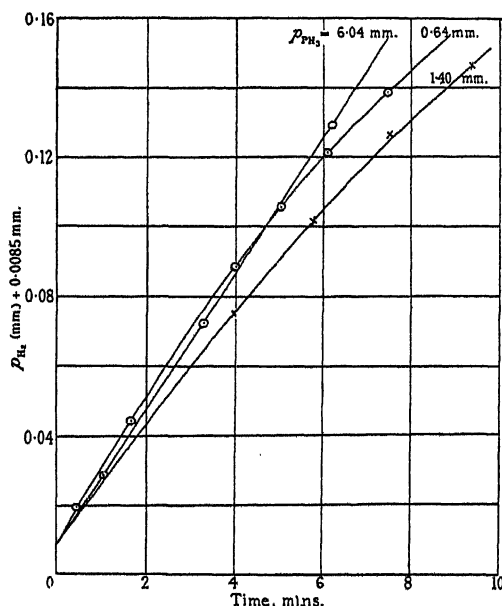


FIG. 2.—The inhibitory effect of hydrogen is shown by the decomposition curves becoming concave to the time axis as p_{H_2} increased. The curvature is greater the smaller the pressure of phosphine. The small variation in initial rate of reaction is due to variation in the intensity of the mercury lamp. 0.0085 mm. is the vapour pressure of PH_3 at -183°C .

From equations (1) and (3)

$$1/R = 1/K \left(1 + \frac{k_2 [\text{H}_2] + k_3}{k_1 [\text{PH}_3]} \right), \quad (4)$$

and therefore if $1/R$ is plotted against $[\text{H}_2]$ for a constant $[\text{PH}_3]$ a straight line should be obtained. Further, if R_0 and R_{H} are the values of R at the beginning of the reaction and after the hydrogen has accumulated, then from (4)

$$\frac{R_0 - R_{\text{H}}}{R_{\text{H}}} = \frac{k_2 [\text{H}_2]}{k_1 [\text{PH}_3] + k_3}, \quad (5)$$

whereby it is then possible to calculate k_2/k_3 if k_1/k_3 is known. The ratio k_1/k_3 is obtained by the method detailed on p. 384.

The horizontal lamp and spiral were used, the phosphine pressure being 4.40 mm. and exposure time 2 minutes. Tables II and III give the record of a typical series of experiments.

Table II.

p_{H_2} added (mm.).	Total p_{H_2} after experiment.	R mm./2 minutes.	Mean p_{H_2} during experiment.
0.000	0.0225	0.0225	0.0112
0.0435	0.0620	0.0185	0.0530
0.0775	0.0935	0.0106	0.0855
0.1080	0.1210	0.0130	0.1145
0.1405	0.1500	0.0095	0.1455

Table III.

p_{H_2} .	R.	k_2/k_3 .	p_{H_2} .	R.	k_2/k_3 .
0.00	0.0233	—	0.08	0.0163	32.5
0.02	0.0216	23.9	0.10	0.0145	36.7
0.04	0.0200	25.0	0.12	0.0125	43.5
0.06	0.0180	29.8	0.14	0.0103	54.5

The results in Table III were obtained by plotting the mean p_{H_2} against R in Table II, and extrapolating back to R_0 . p_{H_2} and R of Table III are then read off from the curve and k_2/k_3 calculated by means of (5). Fig. 3 shows R and $1/R$ plotted against p_{H_2} . As is seen, (a) from fig. 3 the plot of $1/R$ against

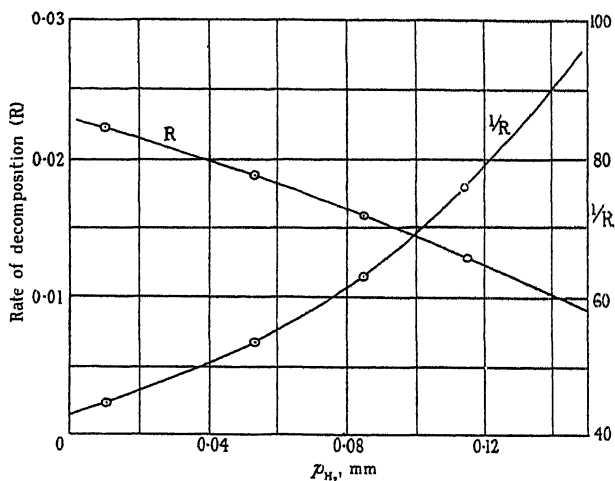


FIG. 3.—The reciprocal of the rate of decomposition of PH_3 is not proportional to p_{H_2} , as would be expected theoretically.

p_{H_2} is not a straight line but a curve convex to the p_{H_2} axis, and (b) from Table III, k_2/k_3 does not remain constant but increases with p_{H_2} . That is, the inhibitory effect of hydrogen is greater than that predicted on the assumption that deactivation alone is responsible for the decrease in the rate of decomposition.

It might be suggested that this additional effect is due to the hydrogen atoms, produced by collisions between excited mercury atoms and H_2 , recombining with the red phosphorus deposited on the walls of the reaction vessel. There are, in the literature two records of the action of atomic hydrogen on yellow phosphorus. Langmuir* found that although a hot (1200° K.) tungsten filament dissociates PH_3 into yellow phosphorus (which is deposited on the walls of the tube) and hydrogen, when the filament was heated to a higher temperature the atomic hydrogen, produced by impact of H_2 on the tungsten surface, is able to recombine with the phosphorus to regenerate phosphine. Similarly Bonhoeffer† found that atomic hydrogen produced in a Wood's tube readily reacts with white phosphorus to give PH_3 , but in addition, a brown iridescent film was formed in the neighbourhood of the phosphorus. It is of importance to note that this film was apparently unattacked by atomic hydrogen. If the brown deposit is similar to that obtained in the photosensitised decomposition of phosphine, then atomic hydrogen ought not to retard the reaction in this manner.

However, experiments were made to find if, *under the experimental conditions employed in this work*, there was any reaction between phosphorus and atomic hydrogen produced photochemically. The method employed consisted in admitting hydrogen to the silica spiral with trap T_3 at 15° , T_2 at -80° , and T_5 at -183° . The lamp was switched on and the normal clean up of hydrogen measured, then T_2 was warmed to 15° so that phosphorus vapour mixed with the hydrogen. If any combination occurred between the phosphorus vapour or the red phosphorus on the spiral and the atomic hydrogen then, owing to T_5 being at -183° , the rate of clean up of the hydrogen should be markedly increased. The pressure of hydrogen is plotted against time of illumination in fig. 4 and it is at once seen that the rate of clean up is unaffected by the presence of phosphorus. As is customary with clean up phenomena the rate decreases as the silica surface becomes covered with atomic hydrogen. The rate of clean up amounts to the right order of magnitude to account for the slight deficiency of hydrogen produced during the decomposition of the phosphine.

These experiments seemed to indicate that recombination of atomic hydrogen occurred with an intermediate product of the dissociation, for example, PH_2

* 'J. Amer. Chem. Soc.,' vol. 34, p. 1310 (1912).

† 'Z. Phys. Chem.,' vol. 113, p. 199 (1924).

or PH radicles. The existence of PH_2 is not known, but apparently PH radicles can exist in flames* and discharge tubes.† Since these radicles probably have only a comparatively short life, it was necessary to devise an experiment whereby atomic hydrogen could be produced in the reaction tube independently of, but simultaneously with the photo-dissociation of the phosphine molecule, and to observe whether the rate of dissociation is decreased.

It is probable that the initial process in the photosensitised decomposition of PH_3 , $\text{Hg}(2^3\text{P}_1) + \text{PH}_3 = \text{PH}_2 + \text{H} + \text{Hg}(1^1\text{S}_0)$ is similar to the direct

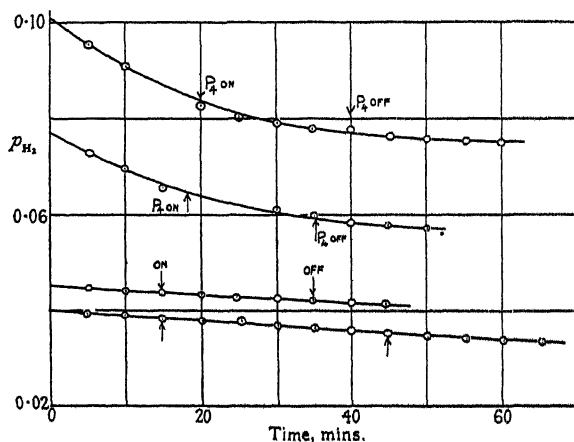


FIG. 4.—The clean up of the hydrogen is not affected by the presence of phosphorus vapour, hence no phosphine is formed.

reaction, $h\nu + \text{PH}_3 = \text{PH}_2 + \text{H}$. On the strength of this assumption, the basis of the idea for testing the above hypothesis is as follows. Phosphine is mixed with a large excess of hydrogen and illuminated, (a) with light from a zinc spark and the amount of decomposition noted, (b) simultaneously with a zinc spark and a water-cooled mercury lamp. In the first case, the excess hydrogen has no effect on the direct reaction; in the second, however, the photosensitised decomposition is inhibited with the simultaneous production of atomic hydrogen which ought to combine with PH_2 or PH and so retard the dissociation brought about by the zinc spark.

Owing to experimental limitations the rate of the photosensitised reaction was quite comparable with the direct reaction. The latter could not be increased because of the low pressure of phosphine and consequent small absorption of radiation; the spark also could not be conveniently increased in

* Ludlam, 'Nature,' vol. 128, p. 271 (1930).

† Pearse, 'Proc. Roy. Soc.,' A, vol. 129, p. 328 (1930).

intensity as it was already consuming 1 amp. at 4000 volts. Separate experiments were therefore made with arc and spark and the total amount of hydrogen produced compared with that from the simultaneous illumination. Inspection of Table IV shows that there is in every case a much smaller amount of hydrogen obtained with simultaneous illumination. In order to prove the point beyond doubt, the two experiments recorded at the bottom of Table IV show that the effect is not observed if hydrogen is absent. A 2-cm. silica tube was used with the arc and spark on either side.

Effect of Tube Diameter.

The next question which arises is whether the recombination occurs at the walls, since owing to the low pressures, ternary collisions, such as $\text{PH}_2 + \text{H} + \text{X} = \text{PH}_3 + \text{X}'$, where X is a third gas molecule, are extremely infrequent compared with collisions with the walls. If the walls do play some essential part in the mechanism of the reaction, then it might reasonably be supposed that the surface/volume ratio of the reaction tube would have some influence on the rate of decomposition. The problem is reduced to finding the value of k_1 of equation (1) either absolutely or relatively to some fixed coefficient, *e.g.*, k_2 .

Table IV.

p_{PH_2} (mm.).	$p_{\text{PH}_2} + p_{\text{H}_2}$ (mm.).	p_{H_2} after exposure.	p_{H_2} produced from PH_3 .	Source of illumination.
0.0865	0.1360	0.0720	0.0225	Arc.
0.0850	0.1345	0.0685	0.0190	Spark.
0.0850	0.1335	0.0810	0.0325	Spark + arc.
0.0575	0.1455	0.0970	0.0050	Arc.
0.0560	0.1465	0.1020	0.0115	Spark.
0.0535	0.1405	0.0935	0.0065	Spark + arc.
0.0555	0.1445	0.0975	0.0085	Spark + arc.
0.0570	0.1470	0.0985	0.0105	Spark.
0.0575	0.1455	0.0930	0.0050	Arc.
0.0555	0.1110	0.0695	0.0140	Arc.
0.0570	0.1110	0.0615	0.0075	Spark.
0.0570	0.1145	0.0735	0.0150	Spark + arc.

Table IV—(continued).

p_{PH_3} (mm.).	$p_{\text{PH}_3} + p_{\text{H}_2}$ (mm.).	p_{H_2} after exposure.	p_{H_2} produced from PH_3 .	Source of illumination.
HYDROGEN ABSENT.				
0.0550	—	—	0.0170	Arc.
0.0570	—	—	0.0120	Spark.
0.0585	—	—	0.0300	Spark + arc.
0.0580	—	—	0.0190	Arc.
0.0585	—	—	0.0120	Spark.
0.0585	—	—	0.0300	Spark + arc.

If $[\text{H}_2] = 0$, equation (4) assumes the simple form

$$1/R = 1/K \left(1 + \frac{k_3}{k_1 [\text{PH}_3]} \right), \quad (6)$$

so that the ratio of the intercept to the slope of the line obtained by plotting $1/R$ against $1/[\text{PH}_3]$ is k_1/k_3 . This equation which indicates how R varies with $[\text{PH}_3]$ will only be valid provided that the pressure of phosphine is low enough to prevent pressure broadening of the absorption line of mercury vapour. This was ensured in these experiments by working with pressures below 10 mm. Tubes of 2, 1, and 0.5 cm. bore were used, the results are given in Tables V–VII and the data in Table VII plotted in fig. 5. Correction is

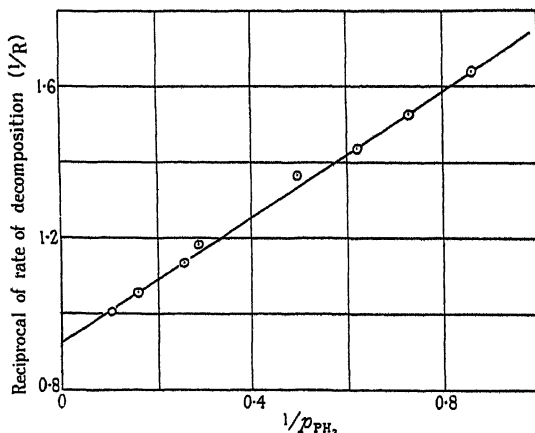


FIG. 5.—The reciprocal of the rate of decomposition of PH_3 is proportional to $1/p_{\text{PH}_3}$, as is to be expected on the simple theory of the effect.

made for the deposition of red phosphorus on the walls. Since the value of k_1/k_3 increases with increasing diameter of tube, it would appear that recombination of the products is facilitated by larger surface/volume ratios and therefore that recombination occurs on the walls.

Table V.—2-cm. Tube.

p_{PH_3} (mm.).	$1/p_{PH_3}$	Rate, (mm./min.)	Corrected rate (R).	1/R.	p_{PH_3} (mm.).	$1/p_{PH_3}$	Rate (mm./min.)	Corrected rate (R).	1/R.
10.00	0.100	0.0425	1.00	1.00	10.00	0.100	0.0380	1.00	1.00
1.26	0.80	0.0300	0.75	1.33	1.80	0.56	0.0280	0.80	1.25
0.62	1.61	0.0245	0.66	1.52	0.56	1.79	0.0220	0.67	1.49
0.36	2.78	0.0170	0.50	2.00	0.32	3.12	0.0160	0.52	1.92
0.20	5.00	0.0140	0.45	2.22	0.20	5.00	0.0105	0.36	2.78
10.00	0.100	0.0285	1.00	1.00	10.00	0.100	0.0270	1.00	1.00
$k_1/k_3 = 2.70.$					$k_1/k_3 = 2.7.$				

Table VI.—1-cm. Tube.

p_{PH_3} (mm.).	$1/p_{PH_3}$	Rate (mm./min.)	Corrected rate (R).	1/R.	p_{PH_3} (mm.).	$1/p_{PH_3}$	Rate (mm./min.)	Corrected rate (R).	1/R.
5.16	0.194	0.0360	1.00	1.00	10.60	0.094	0.0285	1.00	1.00
3.60	0.278	0.0330	0.91	1.10	5.24	0.190	0.0245	0.86	1.16
1.90	0.525	0.0285	0.79	1.27	2.20	0.455	0.0210	0.74	1.35
0.96	1.04	0.0230	0.64	1.56	1.16	0.86	0.0190	0.67	1.49
$k_1/k_3 = 1.60.$					$k_1/k_3 = 1.56.$				

Table VII.—0.5-cm. Tube.

p_{PH_3} (mm.).	$1/p_{PH_3}$	Rate (mm./min.)	Corrected rate (R).	1/R.	p_{PH_3} (mm.).	$1/p_{PH_3}$	Rate (mm./min.)	Corrected rate (R).	1/R.
10.00	0.100	0.0330	1.00	1.00	10.00	0.100	0.0335	1.00	1.00
4.22	0.237	0.0290	0.88	1.14	7.00	0.143	0.0315	0.95	1.05
1.62	0.62	0.0230	0.70	1.43	3.44	0.290	0.0270	0.83	1.20
1.16	0.86	0.0205	0.61	1.64	2.00	0.500	0.0230	0.72	1.39
2.40	0.417	0.0295	0.87	1.15	1.40	0.71	0.0205	0.64	1.56
10.00	0.100	0.0345	1.00	1.00	0.82	1.22	0.0175	0.56	1.79
					10.00	0.100	0.0315	1.00	1.00

$$k_1/k_3 = 1.15.$$

Comparison of the PH₃ and NH₃ Reactions.

It is interesting to compare the phosphine results with those of ammonia obtained by Mitchell and Dickinson* who suggested a mechanism involving excited ammonia molecules. The existence of these molecules is rendered exceedingly improbable by Bonhoeffer and Farkas' work† on the photochemistry of ammonia. If the mechanism suggested in equations (1) and (2) also holds for the decomposition of ammonia, then from Mitchell and Dickinson's results the value of k_1/k_3 may be readily obtained and amounts to 0.20 mm.^{-1} . Whether k_1/k_3 depends on the diameter of the insolation tube is not yet known, but it is evident that the phosphine reaction, according to these calculations, is about ten times as fast as the ammonia decomposition. This result was confirmed by a direct comparison using pressures of ammonia and phosphine of 6 mm. in a cylindrical tube 2 cm. in diameter when 0.00115 mm. of N₂ and H₂ were produced per minute from NH₃ and 0.0118 mm./min. of hydrogen from PH₃.

It is impossible at present to find if this larger yield is due to the greater efficiency of PH₃ as a deactivator of Hg 2³P₁, or to the reactions subsequent to the dissociation of the PH₃ molecule being more efficient than the corresponding ammonia reactions. The effect may be due to both factors.

Another important calculation concerning the ammonia reaction may be made from the results of Mitchell and Dickinson. From equation (5) if $[\text{H}_2]_{\frac{1}{2}}$ is the pressure of hydrogen required to reduce R to $\frac{1}{2}R_0$ then

$$k_2 [\text{H}_2]_{\frac{1}{2}} = k_1 [\text{NH}_3] + k_3. \quad (7)$$

The slope of the line obtained by plotting $[\text{H}_2]$ against $[\text{NH}_3]$ gives k_1/k_2 , i.e., k_1 is obtained relatively to the rate of deactivation of Hg 2³P₁ by H₂. Since the efficiency of this latter process is known absolutely it is then possible to obtain k_1 absolutely. For this purpose the rate of decomposition of NH₃ is plotted against the mean pressure of hydrogen for the interval concerned; extrapolation gives R_0 . The value of $[\text{H}_2]_{\frac{1}{2}}$ is read off and the following results obtained:—

p_{NH_3} (mm.)	p_{H_2} for $\frac{1}{2}R_0$ (mm.)
0.96	0.026
1.88	0.055
3.22	0.077
5.06	0.140

* 'J. Amer. Chem. Soc.,' vol. 49, p. 1487 (1927).

† 'Z. phys. Chem.,' vol. 134, p. 337 (1928).

In calculating their results Mitchell and Dickinson assumed that $[H_2]_i$ was independent of $[NH_3]$ —a supposition which is incorrect as is shown by the table above. From these data $k_1/k_2 = 0.028$. Thus the ammonia decomposition would appear to be much less efficient than quenching by hydrogen. k_1 and k_2 are, however, composite factors and are given by the equation

$$\frac{k_1}{k_2} = \frac{\alpha \sigma_1^2 (1/M_{Hg} + 1/M_{NH_3})^{\frac{1}{2}}}{\sigma_2^2 (1/M_{Hg} + 1/M_{H_2})^{\frac{1}{2}}}, \quad (8)$$

where σ_1 and σ_2 are the *effective* collision radii of the partners Hg and NH_3 and Hg and H_2 , M denotes the molecular weight. α is a factor which denotes the number of ammonia molecules decomposed per collision which is effective in deactivating the excited mercury atom. Zemansky* has recently carried out a series of very careful experiments on the quenching of mercury resonance radiation from which it is possible to obtain values for σ . For NH_3 $\sigma_1^2 = 2.94 \times 10^{-16}$ cm.² and H_2 , $\sigma_2^2 = 6.01 \times 10^{-16}$ cm.². Inserting these values in equation (8) $\alpha = 0.162$. The corresponding result of Warburg† for the direct decomposition, *viz.*, $\gamma = 0.25$ (γ = quantum efficiency) would indicate that, if the collision between Hg 2^3P_1 and NH_3 resulting in deactivation of the mercury atom also results in fission of the NH_3 into NH_2 and H then, in view of the correspondence between γ and α the processes succeeding the elementary decomposition may possibly be identical.

Another interesting point of resemblance of the photosensitised reactions lies in the fact that $1/R$ is plotted against the pressure of hydrogen, curves convex to the abscissæ are obtained as is the case with phosphine. Here again hydrogen would appear to exert an inhibitory effect on the reaction which is greater than that calculated on the assumption of simple deactivation. The explanation is probably the same and, furthermore, the low quantum yield is possibly due to wall recombination of the primary products of dissociation.

As has been shown above the phosphine reaction is about 10 times as fast as the ammonia reaction hence from equation (7) it is at once seen that the inhibition by hydrogen is much less marked. Small pressures of phosphine have to be used in order to obtain accurately measurable effects; this, in turn, introduces the difficulty that the phosphine pressure diminishes during an exposure and with it the reaction velocity. However, extended attempts

* 'Phys. Rev.', vol. 36, p. 919 (1930).

† SitzBer. preuss. Akad. Wiss., p. 741 (1911); compare also Wiig and Kistiakowsky, 'J. Amer. Chem. Soc.', vol. 54, p. 1906 (1932).

were made but the results were so variable that little use could be made of them to calculate k_1/k_2 .

The Photosensitised Oxidation.

Phosphine and oxygen inflame spontaneously between certain limiting pressures,* so that photochemical experiments must be confined to pressures either above or below the explosion limits. The experiments described below are concerned with pressures at or below the lower limit. First of all, it was thought that by adding oxygen to the phosphine the rate of decomposition would be increased since the products of the primary dissociation would be oxidised before they had an opportunity to recombine. The experimental procedure was as follows: phosphine was admitted to the reaction tube and the pressure measured on the McLeod gauge; it was then condensed out with liquid air and oxygen admitted to a known pressure such that the total pressure of PH_3 and O_2 was below the lower limit when the PH_3 evaporated. The exposure was made and the pressure of the gas not condensable by liquid air determined, this gas was then pumped off and the pressure exerted by the residual phosphine (after evaporation) recorded. Small corrections were applied owing to the volatility of the phosphine at -183° . Inspection of Table VIII shows that a considerable increase in reaction velocity is obtained in the presence of oxygen thus lending apparent support for the recombination

Table VIII.

Initial p_{PH_3} .	p_{O_2} added.	p of uncondensed gas after experiment.	p_{PH_3} decomposed.
0.0750	0.0285	0.0340	0.0280
0.0760	—	—	0.0150
0.0885	0.0710	0.0795	0.0605
0.0885	—	—	0.0070
0.0865	0.0220	0.0390	0.0405
0.0885	—	—	0.0070
0.0915	0.0890	0.0670	0.0525
0.0885	—	—	0.0160

hypothesis. The curious anomaly is that the pressure of uncondensable gas is about equal to that of the oxygen. Since oxygen accelerates the reaction

* Dalton and Hinshelwood, 'Proc. Roy. Soc.,' A, vol. 125, p. 294 (1929); Dalton, *ibid.*, A, vol. 128, p. 263 (1930).

it must be used up and, further, any oxide that might be formed is involatile at -183° so that the residual gas will probably be hydrogen. Therefore, the scheme of the reaction may be amended: the oxygen attacks the products of dissociation, forming oxides of such an energy content that they are able to activate phosphine or oxygen molecules which initiate a stable chain reaction between the residual PH_3 and O_2 . If this supposition is correct then the subsequent addition of argon to the $\text{PH}_3\text{-O}_2$ mixture ought to increase the rate still further, precisely in the same way as argon accelerates* the stable chain oxidation of phosphorus. The data given in Table IX show that the surmise is correct. Argon does not accelerate the reaction in absence of oxygen. It is therefore not possible to say whether wall recombination is retarded by oxidation owing to the intervention of the chain mechanism.

Table IX.

Initial p_{PH_3} .	p_{O_2} .	p_{A} .	p uncondensable gas.	p_{PH_3} decomposed.
0.0565	—	—	—	0.0115
0.0500	0.0375	—	0.0275	0.0245
0.0545	0.0425	0.0940	0.1215	0.0350
0.0875	—	—	—	0.0125
0.0875	0.0310	—	0.0385	0.0365
0.0660	0.0325	0.0900	0.1355	0.0455
0.0590	—	—	—	0.0135
0.0610	0.0325	—	0.0420	0.0295
0.0590	0.0325	0.1170	0.1520	0.0435

Effects at the Lower Limit in $\text{PH}_3\text{-O}_2$ Mixtures.

When a $\text{PH}_3\text{-O}_2$ mixture below the lower limit is illuminated by a mercury lamp, Hinshelwood and Clusius† found that on compression explosion took place at a much lower pressure than the normal unilluminated mixture. The effect, however, is not permanent but decays slowly, the original explosion pressure being reached. On further investigation the phosphine proved to be the origin of the effect and Hinshelwood and Clusius suggested that the phosphine molecule became activated by absorption of light; the presence of this activated molecule of long life produces the lowering of the limit. From the decay curves it was estimated that the active molecule could make as

* Melville and Ludlam, 'Proc. Roy. Soc.,' vol. 135, p. 315 (1932).

† 'Proc. Roy. Soc.,' A, vol. 129, p. 589 (1930).

many as 10^5 collisions with the walls before being deactivated. Since no excited or metastable molecule could undergo so many collisions with a wall without becoming deactivated, it was concluded that the molecule was active in virtue of its chemical unsaturation rather than its high energy content.

Now in the theory of the oxidation of phosphine* the concentration $[X_0]$ of one of the chain propagators is given by

$$[X_0] = \frac{k_1 k_2 [O_2] F(c) + k_1 K \cdot F(c)}{(1 - \alpha) k_2 k_3 [PH_3][O_2] + k_2 K [O_2] + k_3 K [PH_3] + K^2} \quad (9)$$

where k_1 , k_2 and k_3 are velocity coefficients and K the rate of deactivation of X_0 at the walls, $F(c)$ is the rate of production of chain centres. The condition for explosion is that $[X_0] = \infty$ or that the denominator of (9) is zero. $F(c)$ does not appear in the denominator hence the explosion pressure ought to be independent of the rate of production of chain centres, therefore, the presence of active molecules produced photochemically should not alter the explosion limit. From (9) then, at the explosion limit

$$[PH_3][O_2] = K' \times \text{constant},$$

K' expressing the efficiency of deactivating collisions of X_0 with the walls. A variation in the condition of the walls may consequently alter the explosion limit.

Before proceeding, the question which required to be settled was whether the effect was due to the walls in spite of Hinshelwood and Clusius' demonstration that illumination of the tube itself did not change the explosion pressure.

For this purpose two similar silica tubes provided with graduated scales and mercury reservoirs were connected together by a short length of tubing having a tap in the middle. This apparatus was connected to sources of PH_3 and O_2 , pumps, etc. Phosphoric acid was used to coat the insides of the illumination tubes in order to maintain constant surface conditions. A hot mercury arc was employed for most experiments, although a water-cooled lamp was also used but owing to the intense resonance radiation it had to be placed at a considerable distance from the tube in order to avoid excessive photo-oxidation. No filters were employed. Let A and B (in which the explosion pressures are determined) be the two tubes. The principle of the method consists in illuminating a PH_3-O_2 mixture in A and transferring immediately afterwards to B where the explosion pressure is measured. If the effect is a surface phenomenon, then the explosion pressure should be identical with that of a similar

* Dalton and Hinshelwood, *loc. cit.*

unilluminated mixture, because the active surface remains behind in A. On the other hand, if the active material is present in the gas it should produce in B a diminished explosion pressure.

Hinshelwood and Clusius' experiments were repeated; for example, after 4 minutes' illumination of B the explosion pressure fell from 19.5 to 17.1 (arbitrary) units. On allowing the illuminated mixture to stand for 4 minutes in B and also in contact with the tubes connecting A and B, the explosion pressure had risen to 17.3. Since the time of transference was about 15 seconds no appreciable loss of active product would occur in this interval.

In one experiment when the mixture was illuminated in A and the explosion pressure determined in B the latter amounted to 22.1 units. Without illumination it was 22.2. In two succeeding experiments the respective values were 20.5 and 20.5; 22.0 and 21.6. There seemed no doubt therefore, that the alteration of surface conditions was responsible for the greater explosibility of the illuminated mixture.

Since the effect has its origin primarily in the phosphine molecule, one explanation would be that the PH_2 radicles or H atoms are adsorbed on the walls which then become less efficient in deactivating the chain propagators, the chain length thereby being increased and the explosion pressure consequently diminished. But it has been seen that in presence of oxygen the rate of decomposition of the phosphine is greatly increased. If this fact is taken into consideration with Frankenburger and Klinkhardt's extensive work* on the mercury photosensitised $\text{H}_2\text{-O}_2$ reaction, which proceeds initially by means of the ternary collision $\text{H} + \text{H}_2 + \text{O}_2 = \text{H}_2\text{O} + \text{OH}$, then it is most probable, in view of the comparatively low pressures (< 1 mm.) and therefore infrequency of ternary collisions, that the phosphorus containing residue of the products of primary decomposition is immediately attacked by oxygen. The hydrogen atom, on the other hand, diffuses to the walls and is most likely adsorbed there.

This hypothesis is easily tested, for it is only necessary to pretreat the surface of tube B with hydrogen atoms produced photochemically by the reaction $\text{Hg } 2^3\text{P}_1 + \text{H}_2$ and determine explosion pressures. The water-cooled lamp was used during these experiments. Table X gives one series of results. The walls of the tube were washed down with phosphoric acid after each experiment and the pressure of hydrogen amounted to about 0.05 mm.

* 'Z. phys. Chem.,' vol. 15, B, p. 569 (1932).

Table X.

Explosion pressure.	Remarks.
19.0	Dark.
16.6	10 min. exposure.
19.6	Dark.
17.8	10 min. exposure.
21.8	Dark.
18.4	10 min. exposure.
20.4	Dark.

There is quite a definite decrease in explosion pressure. Separate experiments showed that the clean up of hydrogen was about 0.004 mm. per 10 minutes, the length and diameter of B being 15 cm. and 2 cm. respectively.

It remains to account for the decay of the effect. This may simply be due to the hydrogen atoms diffusing into the liquid phosphoric acid. Hence the decay period will depend on the phosphoric acid to some extent. Hinshelwood and Clusius actually found that large variations were obtained depending on the previous history of the acid. If the tube is not coated with acid another explanation of the decay can be advanced.

It has been shown by M. C. Johnson* that hydrogen atoms may remain on silica surfaces for long periods of time—much longer than the decay of the above effect. Recombination of the atomic hydrogen would therefore not account for the phenomenon. What would seem more likely is the attack of atomic hydrogen by impact of molecular oxygen on the surface.

Hydrogen saturated with mercury vapour was therefore admitted to a clean silica tube and part of it cleaned up by exposing the tube to the water-cooled lamp. Oxygen was then admitted and the pressures measured at suitable intervals. In Table XI it can be seen that the oxygen slowly attacks the

Table XI.

p_{H_2} initial	0.0735 mm.	0.0660 mm.	0.0465 mm.
p_{H_2} after clean up	0.0695 mm.	0.0625 mm.	0.0445 mm.
$p_{H_2+O_2}$	0.0910 mm.	0.0865 mm.	0.0800 mm.
$p_{H_2+O_2}$	After 8.5 minutes 0.0905 mm.	After 9 minutes 0.0855 mm.	After 10 minutes 0.0780 mm.
$p_{H_2+O_2}$	After 18.5 minutes 0.0900 mm.	After 19 minutes 0.0845 mm.	After 20.5 minutes 0.0780 mm.
$p_{H_2+O_2}$	—	After 29 minutes 0.0845 mm.	—

* 'Proc. Roy. Soc.,' A, vol. 123, p. 603 (1929); *ibid.*, A, vol. 123, p. 447 (1930); 'Proc. Phys. Soc.,' vol. 42, p. 490 (1930).

adsorbed hydrogen, since the total pressure of hydrogen + oxygen falls to a steady value after about 10–20 minutes. A liquid air trap was connected to the apparatus when the oxygen was admitted.

Langmuir* observed a similar disappearance of oxygen when it was admitted to a bulb in which hydrogen atoms had been produced by thermal dissociation on a hot tungsten filament.

In their experiments Hinshelwood and Clusius considered that light in the region 250–280 $\mu\mu$ (transmitted by a chlorine-bromine filter) was responsible, so that the initial stage of the reaction must have consisted in the photo-sensitised decomposition of phosphine due to the presence of some mercury vapour. The effect is not, however, dependent on whether the phosphine molecule is directly or indirectly dissociated so that illumination of the mixture by the unfiltered light from the hot arc would produce the same result.

Discussion of Mechanism of the Decomposition and Oxidation.

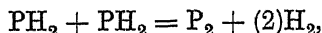
From these experiments the following mechanism can be proposed for the reaction. The phosphine molecule is dissociated, in the first place, by collision with excited mercury atoms into atomic hydrogen and a residue which is probably PH_2 . The products of the dissociation diffuse to the walls of the tube, where they may recombine as is shown by the fact that as the surface/volume ratio of the tube is increased there is a decrease in the rate of decomposition. Recombination is extremely improbable in the gas phase as the concentration of the products of dissociation is small, and since the reaction is accompanied by the evolution of about 100 k. cal. per mole of PH_3 formed, a ternary collision would be necessary to effect stabilisation of the molecule. That recombination can occur between the phosphorus residue and hydrogen atoms is evident from the retardation of the reaction by atomic hydrogen. Formation of PH_3 from red phosphorus or phosphorus vapour and atomic hydrogen could not be detected in this investigation so that this provides further evidence that the atomic hydrogen unites with an intermediate product of the reaction.

If oxygen is added two effects are produced. First, the oxygen attacks the products of dissociation, thus preventing their recombination and therefore accelerating the rate of decomposition; second, the oxide molecule so formed (probably a phosphorus oxide and not H_2O , H_2O_2 or HO) initiates a stable chain reaction between the remaining phosphine and oxygen as is shown by the acceleration of this reaction by argon.

* 'J. Amer. Chem. Soc.,' vol. 34, p. 1310 (1912).

The inhibitory influence of molecular hydrogen is likewise due to two effects: (a) reduction of concentration of excited mercury atoms by deactivation, (b) reaction between atomic hydrogen and products of dissociation at the walls.

The production of molecular hydrogen during the decomposition is due to combination of atomic hydrogen probably at the walls. The PH_2 or PH radicles also may react at the walls,



the P_2 molecules then polymerising to red phosphorus. P_2H_4 might have been formed just at N_2H_4 is produced under suitable conditions in the photo-dissociation of ammonia.* In view of the fact that the hydrogen pressure developed during the decomposition was almost equal to that predicted from the equation $4\text{PH}_3 = \text{P}_4 + 6\text{H}_2$, any formation of P_2H_4 must have been inappreciable. That P_2 molecules do polymerise on striking a glass surface was proved in an experiment carried out in connection with another investigation. In this, a beam of P_2 molecules obtained by passing P_4 molecules through a fine silica jet at 800°C . was directed towards glass targets cooled to various temperatures down to -80°C . After a suitable interval had elapsed a circular spot of red phosphorus had collected on the target.

The importance of surface effects was again emphasised in extending observations on the discovery of Hinshelwood and Clusius that illumination of $\text{PH}_3\text{-O}_2$ mixtures with a mercury arc resulted in a decreased explosion pressure. From the experimental evidence presented above, the dominant influence of hydrogen atoms is demonstrated. It is not at all clear how reaction chains are more efficiently reflected from a surface which is covered with hydrogen atoms. On p. 392 it was recorded that 0.004 mm. of hydrogen occupying about 300 c.c. or 10^{17} atoms adsorbed on 100 cm.² of silica tube sufficed to produce the effect. A saturated layer of H atoms on 100 cm.² amounts to about 10^{17} atoms.†

The author wishes to express his indebtedness to Dr. E. B. Ludlam for his constant advice and encouragement, to the Carnegie Trustees for a scholarship, to Imperial Chemical Industries, Ltd., and the Trustees of the Moray Research Fund of Edinburgh University for grants towards the cost of the apparatus.

* Gedye, 'J. Chem. Soc.,' p. 1158 (1932).

† Johnson, 'Trans. Faraday Soc.,' vol. 28, p. 161 (1932).

Summary.

The mercury photosensitised decomposition of phosphine has been investigated in some detail and the influence of the pressure of PH₃, H₂, O₂ and argon investigated in order to construct a mechanism for the reaction. The kinetics of this reaction are compared with those of ammonia. The ammonia results are recalculated on the basis of a new mechanism and compared with the direct photochemical reaction.

The lowering of the explosion limit of PH₃-O₂ mixtures by illumination with a mercury arc is shown to be due to a surface effect. The phenomenon can be reproduced by pretreatment of the walls with atomic hydrogen. A mechanism for the effect is suggested.

Bands due to the Hydrogen Molecule: The 2p ³Π Bands of Hydrogen.

By IAN SANDEMAN, Ph.D., late Carnegie Research Scholar in the University of St. Andrews.

(Communicated by H. S. Allen, F.R.S.—Received June 10, 1932.)

A set of new bands in hydrogen ending on the hitherto undistinguished 2p ³Π state has recently been described by Richardson and Davidson.* It is no small achievement on the part of these workers to have sorted out these unknown and complicated regularities from the trackless confusion of the many-lined spectrum. While their allocation of the strong low-quantum members cannot be doubted, and their discovery of the characteristic combinations of the 2p ³Π state has disclosed the general form that such bands must assume, there are, however, serious objections to the arrangement of the bands proposed by them. The main objections are as follows:—(1) The upper states are too irregular to be entirely convincing. The phenomenon of "uncoupling" would certainly lead to irregularity, but not quite of so chaotic a description. (2) The lower states are themselves not quite regular enough. The corresponding 3p ³Π state is well established† and its counterpart, the 2p ³Π state, should be about as regular. (3) The combinations between their lines are not quite of the accuracy which might be expected in view of the well

* 'Proc. Roy. Soc.,' A, vol. 131, p. 658 (1931).

† See the writer's work on the Fulcher Bands, 'Proc. Roy. Soc. Edin.,' vol. 49, p. 48 (1929) and that of Richardson and Das, 'Proc. Roy. Soc.,' A, vol. 122 p. 688 (1929).

established trustworthiness of the spectral measurements of Gale, Monk and Lee.* The writer believes the bands to be capable of a simpler arrangement. In this proposed rearrangement most of the strong lines of Richardson and Davidson's $3d^3II \rightarrow 2p^3II$ and $3d^3\Delta \rightarrow 2p^3II$ systems are considered to belong to one single strong band system. The bands of this system shade towards the violet instead of towards the red, and the upper states show considerable regularity. These bands account for nearly all the unallotted strong lines in the orange. The combinations between the lines are of the accuracy that would be expected. Analysis shows the upper state to be a Δ state. In the present paper only these strong bands will be described. Other fainter bands with the same final state exist, some of which have been studied by Richardson and Davidson. These bands require much study before their true form can be ascertained.

The new arrangement is given in Table I. Here the wave numbers of Gale, Monk and Lee are followed by their intensities in brackets. Lines given by Dufour† as showing a Zeeman effect are marked "Z." Lines given by Merton and Barratt‡ as enhanced at high pressures are marked "H," and those enhanced at low pressures "L." A few lines recorded by MacLennan and Shrum§ as strongly enhanced at very low temperatures are marked "+." The intensities as measured by Kapuscinsky and Eymers|| are quoted in italics. Measurements by Johnson¶ of intensities in the molecular spectrum of hydrogen positive rays are given underlined. In the case of the null band recent measurements of the intensities in a strongly condensed discharge by Bay, Finkelnburg and Steiner** are also given. It is of interest that this band, which is not specially intense, came out strongly under the conditions of their experiments, and this fact has been of no small assistance in the correct allocation of lines to this band. To avoid confusion, the last intensities are doubly underlined. A few faint lines taken from the tables of other workers have been used where the lines are missing in Gale, Monk and Lee's tables. These are indicated thus: Merton and Barratt, "M"; Tanaka,†† "T"; Deodhar,‡‡ "D."

* 'Astrophys. J.,' vol. 67, p. 89 (1928).

† 'Ann. Chim. Phys.,' vol. 9, p. 361 (1906) and 'J. Phys. Rad.,' vol. 8, p. 259 (1909).

‡ "Bakerian Lecture," 'Phil. Trans.,' A, vol. 222, p. 369 (1922).

§ 'Trans. Roy. Soc. Canada,' vol. 18, p. 177 (1924).

|| 'Proc. Roy. Soc.,' A, vol. 122, p. 58 (1928).

¶ 'Proc. Roy. Soc.,' A, vol. 114, p. 697 (1927).

** 'Z. phys. Chem.,' vol. 11, p. 351 (1930).

†† 'Proc. Roy. Soc.,' A, vol. 108, p. 592 (1925).

‡‡ 'Proc. Roy. Soc.,' A, vol. 113, p. 420 (1926).

Notation.—In Table I a departure has been made from the usual plan of arranging the lines in band branches. Since there are six branches, and not many lines in each, there seems little to be gained by an arrangement in branches. Instead, the bands are arranged in two columns according as the initial state is a B or A state. The rotational doublets are thus opposite one another. Each triplet of lines with a common initial term is grouped together and the differences $R(j-1) - Q(j)$ and $Q(j) - P(j+1)$ are given in italics. These are the characteristic combination differences of the 2p³Π state. The fact that the lines coming down from the upper B states are stronger than those from the upper A states has already been referred to by Richardson and Davidson who

Table I.
The (0, 0) Band.

R(1)	17572.31	(10)H	34.6	<u>2</u>		R'(1)	17588.00*	(5)	5.7	<u>1</u>	<u>10</u>	
Q(2)	17451.59	(10)H†	30.7		120.72	Q'(2)	17467.13	(6)H	10.0			120.87
P(3)	17271.56	(4)	13.9		180.03	P'(3)	17287.27	(2)	9.3			179.86
R'(2)	17669.81	(2)	3.3	<u>6</u>	179.91	R(2)	17717.06	(3)	5.9	<u>10</u>		180.04
Q'(3)	17489.90	(3)	9.6	<u>10</u>	237.91	Q(3)	17537.02	(4)	9.6	<u>10</u>		237.72
P'(4)	17251.99	(0a)				P(4)	17299.30	(4)H	14.6	<u>10</u>		
R(3)	17765.17	(3)	5.0	<u>10</u>	237.73	R'(3)	†					
Q(4)	17527.44	(4)H	11.1	<u>10</u>	294.25	Q'(4)	17612.63	(1)		<u>4</u>		293.71
P(5)	17233.19	(2h)	10.3	<u>8</u>		P'(5)	17318.92‡	(1)		<u>?</u>		
R'(4)	17855.86	(0)		<u>6</u>	293.63	R(4)	17986.10¶	(0)				294.14
Q'(5)	17562.23§	(3)		<u>8</u>	348.41	Q(5)	17691.96‡	(0)		<u>?</u>		
P'(6)	17213.82	(0)		<u>5</u>								
R(5)	17940.62	(1)		<u>9</u>	347.61							
Q(6)	17593.01	(3)	3.1	<u>10</u>								
P(7)												

* A member of the system $1s3d^1\Sigma \rightarrow 1s2s^1\Sigma$.

† May be overshadowed by the strong line 17850.87 (7) 23.6.

‡ Bay, Finkelnburg and Steiner give new lines, 17318.76 7 and 17691.49 4 which may be P'(5) and Q(5) respectively.

§ The line Q'(5) seems to be about 0.8 cm.⁻¹ too high.

|| May be overshadowed by the line 17202.51 (4) 14.3 7.

¶ A member of the system $1s^1M \rightarrow 1s2s^1\Sigma$.

Table I—(continued).

The (1, 1) Band.

R(1)	17426.25	(10) +	40.3	$\frac{4}{-}$	115.17	R'(1)	17431.57	(6)	17.1	115.01
Q(2)	17311.08	(9) LZ+	37.0	$\frac{4}{-}$	171.31	Q'(2)	17316.56	(5)	15.6	171.63
P(3)	17139.77	(5)	17.6			P'(3)	17144.93	(4)	10.2	
R'(2)	17508.99	(5)	12.1		171.60	R(2)	17529.00	(5) H	12.0	171.36
Q'(3)	17337.39	(3)	11.5		226.38	Q(3)	17357.64	(6) H	15.6	226.86
P'(4)	17111.01*	(2)				P(4)	17130.78	(3a)		
R(3)	17591.22	(4)	5.7		226.92	R'(3)	17633.88	(1)		226.37
Q(4)	17364.30	(5) H	11.7		279.74	Q'(4)	17407.51	(1)		280.39
P(5)	17084.56	(2)	9.8	$\frac{1}{-}$		P'(5)	17127.12	(0)		
R'(4)	17669.81†	(2)	3.3		280.34	R(4)	17739.54	(1)		279.78
Q'(5)	17389.47	(2)			331.27	Q(5)	17459.76	(2)		331.79
P'(6)	17058.20‡	(0)				P(6)	17127.97	(1)		
R(5)	17743.02	(1)			331.81					
Q(6)	17411.21	(5)			330.33					
P(7)	17030.88	(1)								
R'(6)	17809.53	(0)			330.90					
Q'(7)	17428.63	(1)								

* A member of the system $1s3d^1\Pi_a \rightarrow 1s2s^1\Sigma$.

† The same line as (0, 0) R'(2).

‡ A member of the system $^1L \rightarrow 1s2s^1\Sigma$.

Table I—(continued).

The (2, 2) Band.

R(1)	17279.01	(7) LZ+	46.4	<u>4</u>		R'(1)	17280.68	(3) H	17.5	
Q(2)	17169.22	(10) LZ+	37.8	<u>4</u>	109.79	Q'(2)	17171.27	(4)	13.9	109.41
P(3)	17006.43	(10) LZ+	30.9	<u>4</u>	162.79	P'(3)	17007.59	(1)		163.68
R'(2)	17348.12*	(3)	9.4			R(2)	17355.14	(6) HZ	16.2	<u>2</u>
Q'(3)	17184.46	(3)	10.7		163.66	Q(3)	17192.31	(6) H	22.9	<u>4</u>
P'(4)	16969.39	(1)			215.07	P(4)	16976.03	(5) H	17.5	216.28
R(3)	17416.51	(3)			216.32	R'(3)	17433.50	(1)		215.00
Q(4)	17200.19	(1)			265.78	Q'(4)	17218.50‡	(8)	23.0	267.52
P(5)	16934.41	(1)	9.8			P'(5)	16950.98	(0α)		
R'(4)	17481.61	(1)			267.56		§			
Q'(5)	17214.05	(U)			314.29					
P'(6)	16899.76	(0)								
R(5)	17541.45	(2)			316.51					
Q(6)	17224.94	(1)			360.84					
P(7)	16864.10†	(1)								

* Quoted in error in Gale, Monk and Lee's tables as 17347.97.

† A member of the system $1s3d^1I_P \rightarrow 1s2s^1\Sigma$.

‡ This seems to be a coincidence with a strong line of the unallotted spectrum.

§ The following is a possible identification of the next members :—

R(4)	17498.92	(U)			265.73
Q(5)	17233.19	(2h)	10.3		316.62
P(6)	16916.57	(0)			

The initial term difference derived from them would, however, have a somewhat low value.

Table I—(continued).

The (3, 3) Band.

R(1)	17129.89	(8) LZ+	37.0	<u>6</u>	104.48 154.71	R'(1)	*		
Q(2)	17025.41	(6) LZ	25.6	<u>5</u>		Q'(2)	17026.36	(2)	
P(3)	16870.70	(2)	10.5			P'(3)	*		
R'(2)	17186.00	(2)	8.9		155.81 204.21	R(2)	17187.90	(4)	13.7 154.69
Q'(3)	17030.19	(1)				Q(3)	17033.21	(4) HZ	20.2 206.07
P'(4)	16825.98	(1)				P(4)	16827.14	(2)	7.9
R(3)	17239.70†	(2)	9.9			R'(3)	17244.91	(0)	204.26
Q(4)	*					Q'(4)	17040.75	(1)	
P(5)	16781.72‡	(1a)							
Q'(5)	17036.21	(T-) ?							

* R'(1) seems to be overshadowed by the line 17130.87 (3a), a member of the (1, 1) band, P'(3) by P(3), and Q(4) by Q(2).

† A member of the system $1s3d^1\Sigma \rightarrow 1s2s^1\Sigma$.

‡ P(5) seems to be about 0.15 cm^{-1} too high. The wave-number given by Merton and Barratt would yield a better fit, (17068.60).

The (1, 0) Band.

R(1)	19765.11	(0h)	120.76	
Q(2)	19644.35	(0)		

The (2, 1) Band.

R(1)	19495.61	(MO)	115.07	
Q(2)	19380.54	(1)		

The (1, 2) Band.

Q(2) 15099.79 (DO)

The (3, 2) Band.

R(1)	Missing			
Q(2)	19113.35	(00)	5.0	

The (4, 3) Band.

R(1)	18947.44	(1)	104.53	
Q(2)	18842.91	(0)		2.3

show that this peculiarity has a counterpart in the corresponding singlet band complex of hydrogen.

In the nomenclature of the lines, the notation of Richardson and Davidson is followed, in which the letters, P, Q, and R, refer to rotational transitions between strong or "a" states and P', Q' and R' to those between weak or "s" states. The combination principle, $a \nleftrightarrow a$, $s \nleftrightarrow s$, but not $a \nleftrightarrow s$, rules out six of the theoretically possible band branches so that any line of a band is completely specified in this notation without use of subscript letters. The odd-quantum A states and even-quantum B states of both the initial and the final levels belong to the a group, and *vice versa* for the s group. If the initial rotational level be denoted by F'(j), the final rotational level by F''(j) and the part of the emitted frequency of a band line due to electronic and vibrational quantum transitions by ν_0 , the line Q(2) is given by the expression :

$$Q(2) = \nu_0 + F'_B(2) - F''_B(2).$$

Q(2) is thus strictly Q_{BB}(2). There is no ambiguity, however, in the notation, since the lines Q_{AB}(2) and Q_{BA}(2) are forbidden by the above combination principle, and the line Q_{AA}(2) which corresponds to a transition between weak states is specified as Q'(2). To make the notation clearer the formulæ for the lowest-quantum members are set out in Table II.

Table II.—Notation.

$Q(1) = \nu_0 + F'_B(1) - F''_B(1)$	$Q(1) = \nu_0 + F'_A(1) - F''_A(1)$
$P(2) = \nu_0 + F'_B(1) - F''_A(2)$	$P(2) = \nu_0 + F'_A(1) - F''_B(2)$
$R(1) = \nu_0 + F'_B(2) - F''_A(1)$	$R'(1) = \nu_0 + F'_A(2) - F''_B(1)$
$Q(2) = \nu_0 + F'_B(2) - F''_B(2)$	$Q'(2) = \nu_0 + F'_A(2) - F''_A(2)$
$P(3) = \nu_0 + F'_B(2) - F''_A(3)$	$P'(3) = \nu_0 + F'_A(2) - F''_B(3)$
$R'(2) = \nu_0 + F'_B(3) - F''_A(2)$	$R(2) = \nu_0 + F'_A(3) - F''_B(2)$
$Q'(3) = \nu_0 + F'_B(3) - F''_B(3)$	$Q(3) = \nu_0 + F'_A(3) - F''_A(3)$
$P'(4) = \nu_0 + F'_B(3) - F''_A(4)$	$P(4) = \nu_0 + F'_A(3) - F''_B(4)$

In this table expressions for the lines Q(1), Q'(1), P(2) and P'(2) are given, although an exhaustive search has failed to yield these lines. The absence of these lines shows clearly that the initial terms F'_A(1) and F''_B(1) are missing. Hence the initial state must be a Δ state.

An objection to the notation is that the use of dashes to denote transitions between weak states clashes with the established use of a single dash for initial terms and constants. It should be noted, however, that the special use of a single dash here described is confined to the letters P, Q and R. In the writer's view the neatness of the resulting notation outweighs any disadvantage arising from possible confusion to workers on the spectra of other elements in which the combination principle $a \nleftrightarrow a$ and $s \nleftrightarrow s$, does not play an important part.

Properties of the Band Lines.—Table I permits of a more comprehensive survey of the properties of the lines than did the arrangement of Richardson and Davidson since two further bands corresponding to vibrational transitions, $2 \rightarrow 2$ and $3 \rightarrow 3$, have been added. The strong lines R(1) and Q(2) show a tendency to pass from high-pressure to low-pressure enhancement on passing down the group from (0, 0) to (3, 3). The meaning of this peculiarity must be that under low-pressure conditions the molecules with higher vibrational quanta tend to be more abundant in the lowest rotational quantum states. The same two lines R(1) and Q(2) show a smooth variation in intensity from band to band, those of the (2, 2) band being the most intense.* In the positive ray spectrum, however, those of the (3, 3) band are the most intense. The same two lines† also show enhancement at low temperatures. The other band members, notably R(2), Q(3) and P(4) tend to show high-pressure enhancement. The lines of the (0, 0) band in particular are strongly enhanced in the condensed discharge. It should be noted that Bay, Finkelburg and Steiner only give observations for the lines that are definitely enhanced in their condensed discharge. The other lines must be presumed to be of the same intensity as in the normal spectrum. A search has been made without success for a $4 \rightarrow 4$ band. It may be that this band is missing. It cannot be a strong band, hence the correct allocation of lines to it may be a matter of some difficulty.

The regions of the spectrum, in which the bands corresponding to a decrease in the vibrational quantum number of one unit should occur, are singularly bare of unallotted lines. A few faint possible members are given in Table I. The two lines representing the (1, 0) band are those given by Richardson and Davidson.

Term Differences Deducible from the Bands.—In all cases where we have a P line and an R line with a state in common a term difference of the form $\Delta_2 F(j)$, i.e., $F(j+1) - F(j-1)$, can be extracted. These differences are given in Table III for the final terms and in Table IV for the initial terms. The $\Delta_2 F(j)$ in both cases show a considerable measure of regularity. The first and second differences of the $2p\ ^3\Pi\ \Delta_2 F(j)$ vary exactly in the manner of the $3p\ ^3\Pi$ term

* It is interesting to note that the fact that the (0, 0) band is not the most intense is consistent with the fact that change in the moment of inertia during the band transition is greatest for this band, falling off as we pass to the (3, 3) band. (Cf. the steep and yet regular changes in the values of the $\Delta_2 F(j)$ in Table I.) See Birge, 'Bulletin of Nat. Research Council,' No. 57, vol. 11, p. 140 (1926).

† It may be of interest to those who have followed the gradual elucidation of the hydrogen spectrum to know that the strong R(1) and Q(2) lines form the linear series discovered by W. E. Curtis ('Phil. Mag.,' vol. 1, p. 695 (1926)).

Table III.—Final Term Differences.

$$\Delta_2 F''(j) \equiv F''(j+1) - F''(j-1) = R(j-1) - P(j+1).$$

Band.	<i>j</i> .	$\Delta_2 F''_A(j)$.	1st diffs.	2nd diffs.	$\Delta_2 F''_B(j)$.	1st diffs.	2nd diffs.
(0, 0)	2	300.75			300.73		
	3	417.82	117.07		417.76	117.03	
	4	531.98	114.16	-2.91	531.62	113.86	-3.17
	5	642.04	110.06	-4.10			
(1, 1)	2	286.48			286.64		
	3	397.98	111.50		398.22	111.58	
	4	506.66	108.68	-2.82	506.76	108.54	-3.04
	5	611.61	104.95	-3.73	611.57	104.81	-3.73
	6	712.14	100.53	-4.42			
(2, 2)	2	272.58			273.09		
	3	378.73	106.15		379.11	106.02	
	4	482.10	103.37	-2.78	482.52	103.41	-2.61
	5	581.85	99.75	-3.62	582.35	99.83	-3.58
	6	677.35	95.50	-4.25			
(3, 3)	2	259.19					
	3	360.02	100.83				
	4	457.98	97.96	-2.87	361.76		

differences which are known from the Fulcher Bands. The initial term differences also show striking regularity, but of a type which shows that the terms cannot be expressed to a close approximation by a simple quadratic formula. At the foot of Table IV the quantities $\Delta_2 F'(3)$ have been collected. The fact that the differences between these are very nearly constant* seems very strong evidence for the correctness of the arrangement.

Formal Representation of the 2p ³Π Terms.—An interesting feature of these terms is that they can all be expressed to a very close approximation by the simple band formula :

$$F''(j) = Bj(j+1) - Dj^2(j+1)^2,$$

* Since we have the well-known relation $B_0 = B_\infty - \alpha v$ holding between the B constants of each state we should expect the $\Delta_2 F(j)$ which are approximately equal to $2B(2j+1)$ similarly to follow a linear law.

where the B's are consistent with those of the $2p^3\Pi$ states and the D's are small quantities of magnitude about 0.02 cm^{-1} . The B and D constants

Table IV.—Initial Term Differences.

$$\Delta_2 F'(j) \equiv F'(j+1) - F'(j-1) = R(j) - P(j).$$

Band.	j .	$\Delta_2 F'_A(j)$.	1st diffs.	2nd diffs.	$\Delta_2 F'_B(j)$.	1st diffs.	2nd diffs.
(0, 0)	3	563.27	-123.53		493.61	110.26	-6.70
	4	686.80			603.87	103.56	
	5				707.43		
(1, 1)	3	488.95	119.81		451.45	107.35	-7.69
	4	608.76			558.80	99.66	
	5				658.46	92.87	-6.79
	6				751.33		
(2, 2)	3	425.91			410.08	102.14	-7.32
	4				512.22	94.82	
	5				607.04		
(3, 3)	3	374.40			369.00		

Vertical Differences of the $\Delta_2 F'(3)$.

Band.	$\Delta_2 F'_A(3)$.	1st diffs.	2nd diffs.	$\Delta_2 F'_B(3)$.	1st diffs.	2nd diffs.
(0, 0)	563.27	74.32	-11.28	493.61	42.16	-0.79
(1, 1)	488.95			451.45	41.37	
(2, 2)	425.91	51.51		410.08	41.08	-0.29
(3, 3)	374.40			369.00		

obtained by assuming this formula are shown in Table V. Wherever possible these were calculated from the first four $\Delta_2 F(j)$ by the method of least squares. Those calculated for the (1, 1) band give the best fit, all combinations between the lines involving $j = 6$ or less, except one, being expressed with an accuracy within 0.07 cm^{-1} . The B terms of the (2, 2) band give a poorer fit, the residuals of the $\Delta_2 F'_A(j)$ being -0.10 , $+0.15$, $+0.10$ and -0.13 . Here the deviations seem systematic and are probably due to the fact that the formula used is only an approximation.

All the terms calculated in this way are set out in Table VI. In this table the final B terms of the (0, 0) (2, 2) and (3, 3) bands have, however, been calculated from the A terms by using the known combinations between the lines, the (2, 2) B terms because they appeared to deviate systematically from the formula values, and the (0, 0) and (3, 3) B terms because of the meagreness of the data. It is believed that the terms given are accurate, to the modulus of some constant, within 0.1 cm.⁻¹.

A disadvantage of the formula used is that it is extremely sensitive to small changes in the data. Hence it seems best to aim at obtaining close approximations to the terms of low quantum number and to leave the higher terms out of account, until the data from which they are derived are more accurately established.

Table V.—Final B and D Constants.

Band.	B'' _A .	D'' _A .	B'' _B .	D _B ''. [*]
(0, 0)	30.33078	0.0185121	30.34	0.019
(1, 1)	28.88998	0.0176149	28.91413	0.0180296
(2, 2)	27.49243	0.0168646	27.54188	0.0173758
(3, 3)	26.153	0.0168	26.2	—

Table VI.—Final (2p ³Π) Terms.

Band.	(0, 0)		(1, 1)		(2, 2)		(3, 3)	
	F'' _A (j).	F'' _B (j). [*]	F'' _A (j).	F'' _B (j).	F'' _A (j).	F'' _B (j). [*]	F'' _A (j).	F'' _B (j). [*]
1	60.59	60.45	57.71	57.76	54.92	54.94	52.24	52.35
2	181.32	181.29	172.71	172.84	164.35	164.69	156.32	156.72
3	361.30	361.26	344.14	344.37	327.48	328.04	311.42	312.12
4	599.21	599.08?	570.75	571.07	543.10	543.80	516.34	517.49
5	893.26	—	850.85	851.20	809.60	810.64	—	—
6	1241.24	—	1182.31	1182.59	1124.93	1126.11	—	—

* Obtained from the A terms by means of the combination differences F''_A(j + 1) - F''_B(j) and F''_B(j + 1) - F''_A(j).

The Initial Levels.—One advantage of obtaining values for the final terms is that we can by adding the appropriate final term to each band line eliminate the variable part of the final states and so obtain values for the successive differences between the initial terms.

Thus

$$\begin{aligned}
 \nu_0 + F'_B(2)^* &= R(1) + F''_A(1), \\
 &= Q(2) + F''_B(2), \\
 &= P(3) + F''_A(3), \\
 \nu_0 + F'_B(3)^* &= R'(2) + F''_A(2), \\
 &= Q'(3) + F''_B(3), \\
 &= P'(3) + F''_A(4),
 \end{aligned}$$

and so on. In general there are three data for the determination of each value $\nu_0 + F'(j)$, hence these values are believed to be considerably accurate. The quantities together with the initial term differences $F'(j+1) - F'(j)$ derived from them are shown in Table VII. The first and second differences between $F'_B(j+1) - F'_B(j)$ indicate that these at least follow some regular law. The

Table VII.
Differences between successive Initial A Levels.

Band.	j .	$\nu_0 + F'_A(j)$.	$F'_A(j+1) - F'_A(j)$.		
(0, 0)	2	17648.48	249.87		
	3	17898.35	313.47	63.60	
	4	18211.82	373.38	59.91	-3.69
	5	18585.20			
(1, 1)	2	17489.30	212.52		
	3	17701.82	276.46	63.94	
	4	17978.28	332.35	55.89	-8.07
	5	18310.63			
(2, 2)	2	17335.62	184.20		
	3	17519.82	241.77	57.57	
	4	17761.59			
(3, 3)	2	17182.68	161.95		
	3	17344.63	211.43	49.48	
	4	17556.06			

* These quantities may also contain an unlimited constant part of the final term. This does not, however, affect the validity of the method for determining the

$$F'(j+1) - F'(j).$$

Table VII—(continued).
Differences between successive Initial B Levels.

Band.	<i>j</i> .	$\nu_0 + F'_B(j)$.	$F'_B(j+1) - F'_B(j)$.		
(0, 0)	2	17632.89	218.27		
	3	17851.16	275.32	57.05	—3.78
	4	181126.48	328.59	53.27	—3.05
	5	18455.07	378.81	50.22	
	6	18833.88			
(1, 1)	2	17483.93	197.81		
	3	17681.74	253.64	55.83	—4.27
	4	17935.38	305.20	51.56	—3.50
	5	18240.58	353.26	48.06	—3.32
	6	18593.84	398.00	44.74	
	7	18991.84			
(2, 2)	2	17333.92	178.57		
	3	17512.49	231.51	52.94	—3.76
	4	17744.00	280.69	49.18	—3.51
	5	18024.69	326.36	45.67	
	6	18351.05			
(3, 3)	2	17182.13	160.19		
	3	17342.32	208.80	48.61	
	4	17551.12			

$F'_A(j+1) - F'_A(j)$ also vary in a regular manner except that there seems to be an irregularity in the (1, 1) band.

As a first step towards investigating the form of the initial terms the $\nu_0 + F'_B(j)$ have been expressed by a simple quartic formula:—

$$\nu_0 + F'_B(j) = a + bj + cj^2 + dj^3 + ej^4,$$

where *a*, *b*, *c*, *d* and *e* are constants. For the $\nu_0 + F'_B(j)$ of the (0, 0) band the values found for these constants are as follows:—

$$a = 17386.27,$$

$$b = 54.5808,$$

$$c = 36.35458,$$

$$d = -1.055833,$$

$$e = 0.0304167.$$

The fact that the b and c constants are in the ratio $3/2$ seems significant. It shows that the major part of the term cannot be expressed in the form $Bj(j+1)$. If we could replace this expression by $B(j+\frac{1}{2})(j+1)$ and simultaneously allow an appreciable correction term in j^3 , a good fit could be obtained. Until some theoretical expression for these terms can be found, it seems useless to indulge in speculation.

The Vibrational Terms.—The few of these that are deducible from the bands are shown in Table VIII. In the absence of a definite formal representation of the initial states the null-line formula cannot be calculated.

A rough calculation of this formula may, however, be made by taking the $Q(2)$ lines to represent the null-lines. If we take the usual form of the null-line formula :

$$\nu_{v''v'} = \omega_0'v' - \omega_0'x'v' - \omega_0''v'' + \omega_0''x''v'',$$

v' and v'' being the initial and final vibration quantum numbers ($\frac{1}{2}$ integral) respectively, we obtain :

$$\omega_0' = 2316$$

$$\omega_0'' = 2465$$

$$\omega_0'x' = 62$$

$$\omega_0''x'' = 61.$$

This result is seen to be inconsistent with the rule of Mecke* whereby

$$\frac{B_0'}{\omega_0'} = \frac{B_0''}{\omega_0''},$$

B_0 being the B constant for the null band. B_0'' from Table III is 30.33 and B_0' appears from the formula just calculated to be 36.35 .

The band system thus seems to represent a unique case of one where $\omega_0'' > \omega_0'$, and yet the individual bands shade towards the violet. It must not, however, be forgotten that the formula of Mecke, as Birge* has shown, is qualitative rather than quantitative. And the violet shading becomes much less pronounced as we pass from the $(0, 0)$ to the $(3, 3)$ band: in fact, were further members known, about the $(4, 4)$ band the shading would turn over and point in the opposite direction.

One possibility that must not altogether be put aside is that the lines identified as coming from the initial A states may in reality belong to a different system from those identified as coming from the initial B states. The writer thinks this very unlikely, in view of the regular way in which the doublet

* See Birge, *loc. cit.*, p. 233.

spacing varies from band to band. This possibility would not in any case remove the difficulty about the violet shading of the bands.

The well-marked final term differences which agree with band theory and resemble the 3p ³Π differences provide a strong reason for thinking that the null band has been correctly identified as such.

Table VIII.—The Vibrational Differences.

Final.		
Interval	R(1)/s.	Q(2)/s.
(1, 0) — (1, 1)	2338·86	2333·27
(2, 1) — (2, 2)	2216·60	2211·32
(1, 1) — (1, 2)	—	2211·29
(3, 2) — (3, 3)	—	2087·94
Initial.		
(1, 0) — (0, 0)	2192·80	2192·76
(2, 1) — (1, 1)	2069·36	2069·46
(2, 2) — (1, 2)	—	2069·43
(3, 2) — (2, 2)	—	1944·13
(4, 3) — (3, 3)	1817·55	1817·50

The fact, that only the bands corresponding to a zero transition of the vibration quantum number appear strongly, is not inconsistent with theory, and represents the limiting case when the arms of the parabola of maximum intensity in the intensity diagram approach one another closely.*

An application of the semi-empirical formula of Birge† provides another test for the correctness of the band system. This formula stated in symbols is :

$$R = 2\pi B_0/\alpha.$$

R should be a constant greater than 1, and generally lying between 1·2 and 2·2. Here α is the small quantity used in the formula just quoted for the null lines, and α is the parameter of the B constants which are related to one

* See the work of Condon, 'Phys. Rev.', vol. 28, p. 1182 (1926) and other papers.

† See Loomis and Wood, *ibid.*, vol. 28, p. 1182 (1926).

another by the expression :

$$B_v = B_0 - \alpha v,$$

v being again the vibration quantum number. The result of an application of this formula to the present band system as well as to the Fulcher bands is shown in Table IX.

Table IX.—Comparison of Constants.

	B_0 .	α .	α/B_0 .	ω_0 .	$\omega_0\alpha$.	R.
Fulcher bands $\begin{cases} 2s\ ^3\Sigma & \dots\dots\dots \\ 3p\ ^3\Pi_B & \dots\dots\dots \end{cases}$	$\begin{cases} 33.4 \\ 29.8 \end{cases}$	$\begin{cases} 1.6 \\ 1.4 \end{cases}$	$\begin{cases} 0.048 \\ 0.047 \end{cases}$	$\begin{cases} 2594 \\ 2308 \end{cases}$	$\begin{cases} 0.027 \\ 0.028 \end{cases}$	$\begin{cases} 1.1 \\ 1.2 \end{cases}$
Present bands $\begin{cases} 2p\ ^3\Pi_B & \dots\dots\dots \\ 3d\ ^3\Delta_B & \dots\dots\dots \end{cases}$	$\begin{cases} 30.3 \\ 36.4 \end{cases}$	$\begin{cases} 1.5 \\ 3.0^* \end{cases}$	$\begin{cases} 0.048 \\ 0.09 \end{cases}$	$\begin{cases} 2465 \\ 2316 \end{cases}$	$\begin{cases} 0.025 \\ 0.027 \end{cases}$	$\begin{cases} 1.0 \\ 0.7 \end{cases}$

* Obtained by assuming the B constant to be in an approximately constant ratio to the lowest term difference.

Here again the results vary somewhat from the established experience of band spectra, but not very much more for the present bands than for the well-established Fulcher bands.*

Table IX shows the constants used in the calculation of R, and is, on the whole, very favourable to the new Δ state, except for the high value of B_0 , since there is a reasonable similarity between the constants of the new states and those of the Fulcher states. The spectrum of hydrogen, which holds a limiting position among the elements, need not be expected to follow rigidly the semi-empirical results obtained from statistical observation of other elements.

My thanks are due to Professor H. S. Allen for helpful encouragement.

Summary.

It is suggested that in the tentative $3d\ ^3\Pi_{a,b} \rightarrow 2p\ ^3\Pi$ and $3d\ ^3\Delta_{a,b} \rightarrow 2p\ ^3\Pi$ bands, outlined a year ago by Richardson and Davidson, the weak lines, due to transitions between the antisymmetric states, are incorrect. In $3d\ ^3\Delta_a$ ($v=0$) they should be replaced by the corresponding symmetric lines of

* It may be that a knowledge of the term form for the new state would disclose the reason for the apparently very low value of R (0.7). The upper state of the $2p\ ^3\Sigma \rightarrow 2s\ ^3\Sigma$ gives the value 1.1 in agreement with the remaining states of Table IX. See the writer's work on these bands, 'Proc. Roy. Soc. Edin.,' vol. 49, p. 253 (1929), where, however, the formula for the B_v has been wrongly quoted as $B_v = B_0(1 - \alpha v)$ instead of $B_v = B_0 - \alpha v$.

$3d^3 \Delta_6$; while $3d^3 \Delta_6$ itself is to be formed in the same way from $3d^3 \Pi_6$ and $3d^3 \Pi_6$.

The strong $v = 0$ and $v = 1$ levels of the lower state are as in the original bands. Higher rotational and vibrational levels are added, both in the upper and lower states. The arrangement eliminates the weak points of the original scheme.

The Structure of Surface Films. Part XVI.—Surface Potential Measurements on Fatty Acids on Dilute Hydrochloric Acid.

By N. K. ADAM and J. B. HARDING.

(Communicated by F. G. Donnan, F.R.S.—Received June 11, 1932.)

Guyot,* Frumkin,† and Schulman and Rideal‡ have shown that it is possible, by means of an air electrode covered with a small amount of a radioactive deposit, which ionises the air in its neighbourhood, to measure changes in the contact potential at an air-liquid interface caused by spreading a film over the surface. It is now clear that this change in contact potential is caused by the dipoles of the film molecules, the magnitude of the change in potential depending on the vertical component of the dipole moment of the molecules in the film, and on the extent to which the water molecules and the ions in the solution are re-arranged near the surface under the influence of these dipoles.

In combination with surface pressure measurements, which have already given a great deal of information as to the orientation of the molecules in the surface, and their shapes, sizes, and adhesive fields of force, this method, which indicates the orientation of the dipoles of the film molecules to the surface, is a valuable addition to our methods of investigating the structure of surface films.

The electrical part of the present apparatus follows the general lines of that constructed by Schulman and Rideal. It is combined with the instrument of Adam and Jessop§ for measuring surface pressure on the same film, a combination which is extremely useful in avoiding errors due to collapse of the film.

* 'Ann. Physique,' vol. 10, p. 506 (1924).

† 'Z. phys. Chem.,' vol. 116, p. 486 (1925).

‡ 'Proc. Roy. Soc.,' A, vol. 130, p. 259 (1931).

§ 'Proc. Roy. Soc.,' A, vol. 110, p. 423 (1926). Minor modifications in this instrument are incorporated in the description in "The Physics and Chemistry of Surfaces," Clarendon Press, 1930, pp. 38–42, which describes exactly the instrument now in use.

or leakage past the barriers, which we have found may take place unobserved unless the surface pressure is recorded. The trough was the same size as that used in earlier measurements of surface pressure, $60 \times 14.0 \times 1.8$ cm., and was of brass, waxed. Glass strips, covered with paraffin wax in the usual manner, were used to manipulate the film and to clean the surface. The complication of the apparatus arises from the fact that it is necessary, in order to avoid disturbance from stray electrical charges, to enclose the whole apparatus in an earthed cage, making the necessary adjustments from outside.

The electrical circuit for measuring the potential is shown in fig. 1. AB is the trough, supported on levelling screws standing on insulated plates. C is a nonpolarisable electrode connecting the liquid in the trough to a standard cell G and a potentiometer LSMRH. The slide wire of the potentiometer is

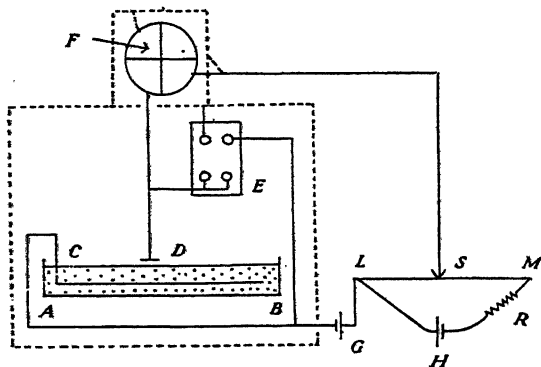


FIG. 1.—Earthed cage shown by dotted line.

earthed at S. One pair, F, of quadrants of the Dolezalek electrometer, which had a very fine phosphor-bronze suspension, were also earthed. The very carefully insulated air electrode D, consisting of a copper disc about 2 cm. diameter, coated with a polonium preparation (kindly given us by Dr. L. F. Bates), was connected with the other pair of quadrants. A silica rod was used to insulate the air electrode. The paraffin block E, which has four small mercury cups, provided a means of earthing both pairs of quadrants, or of connecting one pair to the potentiometer directly, thus permitting calibration of the electrometer deflections directly in millivolts. The needle was charged to about 100 volts by a dry battery, the sensitivity being about 600 divisions on the scale per volt. The trough, electrometer, air electrode, and the leads to the electrometer, were enclosed in an earthed cage of tinfoil on a wood framework, with a removable front of copper gauze: the latter was transparent enough to permit the spot of light from the instrument for measuring surface pressure to be read on the scale Q.

The general arrangement is shown in fig. 2. AB is the trough. One barrier is shown at B (four, two each side of the instrument, were actually used). The instrument for measuring surface pressure is at P. The light falling on the mirror of this instrument was from an unfrosted filament lamp outside the cage, passed through a lens L' in the wall of the cage, and fell on a scale Q, which could be adjusted by rack and pinion from outside the cage, thus permitting instant setting of the scale at any desired zero. The mechanism for moving the barriers from outside the cage was similar to the usual rider mechanism on a balance; the barriers could be lifted off the trough, or moved along it without lifting, by a pair of forked supports N, attached to a carriage (not

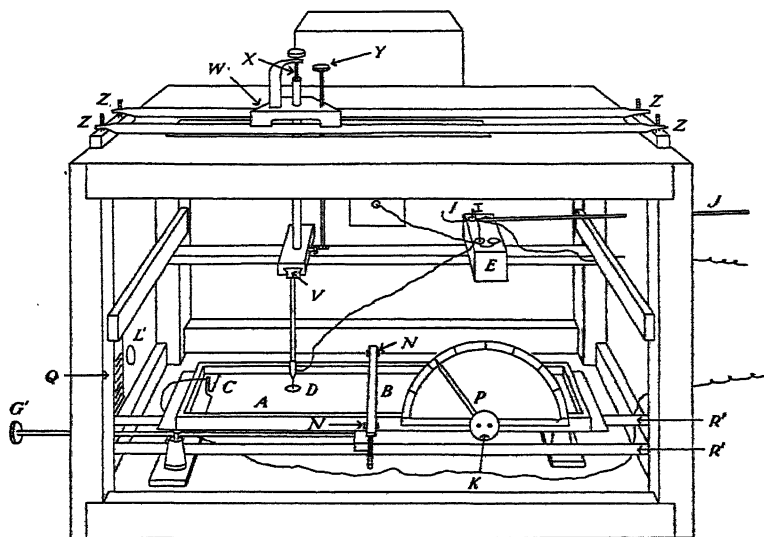


FIG. 2.

shown) which moved on two steel rods R' . This carriage could be moved up and down the trough by means of a long rod, whose head G' projected outside the cage. By turning this rod, a rack and pinion mechanism in the carriage was worked so as to lift the forked supports and the barrier. The usual operations of sweeping the surface, and compressing the film, can thus be accomplished without interfering with the electrostatic shielding.

The torsion heads of the surface pressure instrument, of which one dial K, only, is shown, were controlled from outside the cage by a detachable ebonite handle, on the end of which are two short prongs, which fit into holes on the torsion heads. A small door in the front of the cage was opened when it was necessary to insert this handle.

The air electrode D could be moved the whole length of the trough, and nearly the whole width also, in the following way. The polonium coated disc was soldered to a copper wire held in a chuck fixed to a silica rod, which was fixed to a slide V. This slide could be moved across the trough by rack and pinion worked by a knob Y outside the cage. The slide V was carried on a stout rod, capable of being raised or lowered by a screw X. This rod passed through a carriage W, which could move up and down the length of the cage, with three point contact, on two steel rods ZZ. These rods could be adjusted with leveling screws at each end, so that the air electrode travelled exactly parallel to the surface of the liquid in the trough. Mr. G. E. Fensome designed and constructed several of these devices for controlling the movements from outside the cage.

The wire from the air electrode is connected to one mercury pool in the paraffin block E, from which another wire runs, through a silica tube mounted in a block of paraffin wax in the side of the cage, to the electrometer. A small bridge of copper wire, I, held in a glass handle, J, could be inserted in the left-hand pair of mercury cups, thus earthing the air electrode and both pairs of quadrants and fixing the zero of the electrometer; or in the right-hand pair of cups, connecting the electrometer directly to the potentiometer, allowing calibration of the deflections of the electrometer in millivolts directly. With I removed, the electrometer was connected to the air electrode alone.

The amount of fatty acid in the film spread on the surface was measured by expelling a known quantity of a standard solution of the acid (in petroleum ether 60° – 70° , or a mixture of benzene, 5 parts, alcohol, 2 parts by volume) from a capillary tube mounted on a scale, drawn out to a narrow tip so that small drops could be formed.* The tube was carefully calibrated by means of a mercury thread; the movement of the column of liquid in this tube was controlled by a small screw plunger, connected to one end of the tube by rubber tubing; the whole was mounted rigidly in a stand. In order to place a known amount of material on the surface, the end of the tube was inserted through a small door in the gauze shielding on the front of the cage, the plunger was screwed in a short distance, and the movement of the end of the liquid column read off the scale.

The procedure during an experiment was as follows :—

- (a) The electrometer quadrants were all earthed and the zero of the electrometer taken.

* We are indebted to Dr. A. H. Hughes for pointing out the improved accuracy obtainable by controlled expulsion of a column of liquid from a calibrated tube.

- (b) The surface was cleaned by sweeping several times with the barriers, the quadrants remaining earthed.
- (c) The air electrode was lowered to within 2 or 3 mm. of the surface.
- (d) The pair of quadrants connected to the air electrode were unearthed and the potential difference between the points D and L in the circuit (fig. 1) balanced out by the potentiometer, bringing the electrometer back to zero.
- (e) The film was spread on the surface and the alteration in potential determined either by balancing it out with the potentiometer, or by noting the electrometer deflection; both methods were found to give the same result, and they were frequently combined. This alteration in potential is ΔV .

The time required from the last cleaning of the surface to the first measurement was usually about $1\frac{1}{2}$ minutes.

Tests of residual contamination on the surface after cleaning were frequently made, by reducing the area of the surface to a small amount, and noting the surface pressure set up or the change in potential. A twenty-fold reduction in area usually produced (with freshly cleaned surfaces) a rise in potential of 10–15 millivolts, and a surface pressure less than 0.3 dynes. Surfaces showing much less contamination than this could not be obtained, and the apparatus was usually dismantled and cleaned if greater residual effects than twice the above amounts were found.

The sensitivity of the electrometer deflections were determined two or three times a day; this was necessary as this sensitivity tended to vary by about 2 per cent. over long periods.

The choice of a suitable non-polarisable electrode for connecting the liquid in the trough to the potentiometer presented some difficulty. A calomel cell at one end of the trough was used for a time, and also a silver wire coated with silver chloride. The use of a small electrode at the end of the trough introduced an unexpected complication; when the air electrode was moved about so as to explore a carefully cleaned surface, irregular variations in the potential were found. These were eventually traced to varying contact potential between the brass trough and the liquid; apparently the current from the small non-polarisable electrode short-circuited through the liquid to the brass, then from the brass just underneath the air electrode back to the liquid. These irregularities were overcome by using a large sheet of silver, C, insulated from the trough by an ebonite frame, covering nearly the whole of the bottom of the trough

with a projection above the water at one end; the surface of this silver sheet was coated with silver chloride by anodic treatment in a chloride solution. Variations in the potential from point to point over a clean surface were thus reduced to 5 millivolts or less; and as these remained constant for some time, the values of the change in surface potential produced by the film (ΔV) could be corrected for these zero fluctuations by subtracting from the potentials observed for the film-covered surface at any point, the value obtained at this point in a blank experiment on a clean surface. With a trough of glass or silica, such as those used by Schulman and Rideal, this difficulty should not arise; but a brass trough presents decided advantages if the measurement of surface pressure is required also.

The height of the air electrode above the liquid was not important, identical values of the *change* in potential (ΔV) caused by the film being obtained at any height between 0.5 and 5 mm. from the surface. A variation in the measured potential of the surfaces, both clean and film-covered, up to 5 millivolts, was, however, usually observed on raising the electrode by this amount. It was shown that this was not due to defective insulation, by applying a known electromotive force to the liquid and measuring it on the electrometer through the air gap; the results so obtained were completely independent of the height of the electrode, between 0.5 and 5 mm. A height of 2 to 3 mm. was usually adopted.

Experimental Results.

All the films described in this paper altered the potential positively, *i.e.*, they changed the double layer at the surface so that the outside layer was more positive to the inside than with a clean surface.

The acids used were myristic, palmitic, stearic, and behenic (14, 16, 18, and 22 carbon atoms): Mr. Danielli kindly purified the myristic acid exceptionally well, fractionating the methyl ester at about 15 mm. with a 6-inch column, the fraction taken distilling within 1° ; the ester was hydrolysed and the acid similarly fractionated within 1° , and finally crystallised. The other acids had been crystallised several times and had the correct melting point within $\frac{1}{2}^\circ$; titration indicated the correct molecular weight within 1 part in 200.

Myristic acid has the advantage that condensed, expanded, and vapour films can all be obtained at room temperature, with the transition regions between them. The results for areas between 22 and 60 sq. A. are shown in fig. 3; the two lowest curves show the surface pressure at two temperatures; the next curve above shows the surface potential. This curve is the mean of a

large number of experiments, all of which gave curves almost parallel, but differing slightly in position. The extreme limits of variation between different experiments amounted to about 20 millivolts at 40 sq. A. per molecule, the lowest and highest readings obtained being 198 and 218 millivolts; the majority lay between 203 and 212 millivolts, at this area.

These curves agree fairly well with those in fig. 4 of Schulman and Rideal's first paper,* between 47 and 40 sq. A., and also at lower areas with the curve which they called "metastable." Careful examination of a number of films has, however, failed to show any sign of discontinuity in the potential curve,

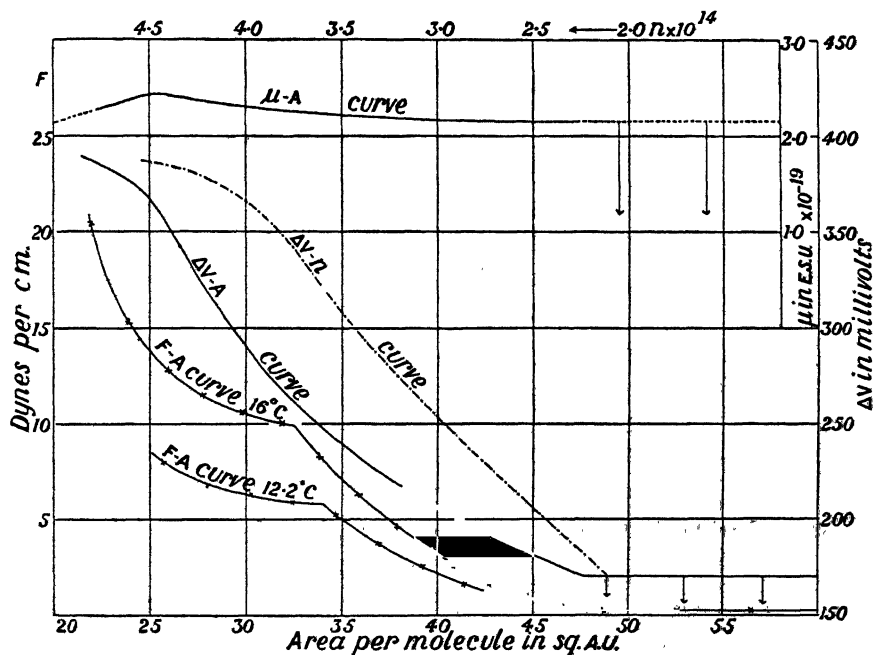


FIG. 3.—Myristic Acid on 0.01 N HCl.

at the point where the sharp change takes place in the pressure curve, from the expanded film to the transitional region to the condensed state. We have obtained only the "metastable" curve of Schulman and Rideal, even after waiting three-quarters of an hour; there is no doubt from our surface pressure curves that the films were actually changing from the expanded to the condensed states. The only occasions on which lower potentials than these could be found were one or two when the surface pressure instrument showed that the film was leaking badly past the barriers or collapsing.

* 'Proc. Roy. Soc.,' A, vol. 130, p. 268 (1931).

Our results show the potential curve ending at the same point at its lower end as the pressure curve, at an area of 47.5 sq. A. with a probable error of 1 sq. A. This agrees well with Adam and Jessop's value* deduced from surface pressure measurements, of 48 ± 2 sq. A., less well with Schulman and Rideal's value of 53.5. The present values seem to us the most reliable of the two former, as they were obtained with the calibrated tube, the earlier results being obtained with a dropping pipette.

The surface has been examined for inhomogeneity by moving the electrode over the surface. In the expanded region, and in the region of transition to the condensed state, i.e., between 47 and 24 sq. A., variations in ΔV greater

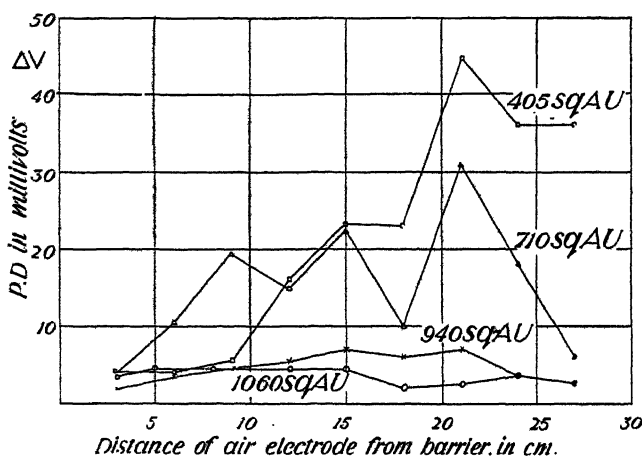


Fig. 4.—Myristic Acid on 0.01 N HCl. Variation of ΔV along trough at different areas.

than about 10 millivolts were not found if the search was rapidly conducted (about 10 minutes). If, however, the search was continued over periods of about 25 minutes, irregularities frequently appeared, their maximum value being about 30 millivolts. It was suspected that these irregularities were due to some collapse of the monomolecular film into aggregates, because the surface pressure had a slight tendency to fall off. Films were therefore examined by the ultramicroscopic technique of Zocher and Stiebel,† in which a cardioid condenser inserted through a hole in the bottom of the trough is focussed on to the surface, so that aggregates show up as bright spots on a dark background. At 35 sq. A. per molecule, practically no aggregates were found during the

* 'Proc. Roy. Soc.,' A, vol. 112, p. 366 (1926).

† 'Z. phys. Chem.,' A, vol. 147, p. 401 (1930). Some minor modifications were introduced, which will be described in a later paper.

first 3 minutes, but after this time a number appeared, indicating that the film was slowly collapsing. These fluctuations appeared more rapidly and were larger at lower temperatures (*ca.* 13°) than at higher (*ca.* 20°); this also is in accord with the theory that the fluctuations are due to the collapse of the films locally, since Cary and Rideal* have shown that a crystalline aggregate has a greater tendency to spread at higher than at lower temperatures, so that the tendency to collapse should be less at higher temperatures. An attempt was made to obtain further evidence in support of this theory by examining whether there was any fluctuation in a freshly spread film, on which a crystal of the solid myristic acid rested, when the air electrode passed over the crystal. No such fluctuation was found; it is, however, probable that the crystal, being more perfectly formed than the aggregate of molecules hastily expelled from the film by collapse, does not affect the surface potential, whereas the less regularly built collapsed aggregate might do so. We think that these slight fluctuations are probably, though not certainly, due to collapse, and not to any inhomogeneity in the film itself. The fluctuations occur equally in the region of pure expanded film and in the transition region from expanded to condensed film, as shown by the surface pressure, and are not therefore due to mixture of separate surface phases, which might be identified with condensed or expanded films. In fact, no evidence has yet been obtained that the "condensed" and "liquid expanded" types of films are really two distinct surface phases.

Transition to Vapour Film.

These curves for the expanded films continue smoothly down to 47.5 sq. A., ΔV falling off continuously with increasing area. At this area, which appears to be definite within about 1 sq. A., the decrease in ΔV stops abruptly. At greater areas violent fluctuations in ΔV appear; if the area is not much greater than 47.5 sq. A., the greater part of the surface is covered with a film having a potential of about 170 millivolts, and with a stationary electrode the film may appear to have this potential up to areas of 65 or 70 or even more. Schulman and Rideal gave the surface potential as about 155 millivolts over this region, and constant up to 88 sq. A. At that time, however, they used a fixed electrode. We have found, in preliminary experiments, that a fixed electrode may give this steady value, or may give a variable value; the variable value is always obtained if the film is gently blown about the surface even

* 'Proc. Roy. Soc.,' A, vol. 109, p. 301 (1925).*

with a fixed electrode. After consultation with Dr. Schulman, we agreed that a moving electrode is the most satisfactory method of examining the surfaces, and we understand he is now in complete agreement with us as to the inhomogeneous nature of the surface in this region. The "homalic" and "pre-homalic" phases are clearly non-existent, since the surface is a two-phase system in the region where these were described.

Our value of 170 millivolts does not agree exactly with Schulman and Rideal's of about 155 millivolts, at the lower end of the expanded curve. We think this slight discrepancy, which is not at present of serious theoretical importance, may be due to different specimens of myristic acid having been used. In our earliest experiments, we obtained values of about 160 millivolts, but more recently, using the most carefully purified specimen described above, we have obtained higher values. We understand that, more recently, Drs. Hughes and Schulman have also *sometimes* obtained higher values, about 170 millivolts, using our new specimen of the acid. There is still a slight uncertainty in this value. We have found values from 162 to 178 millivolts, on the best specimen of the acid, without being able to account for the variation in the values. These are decidedly larger than the experimental error, which we estimate as about 4 millivolts in our experiments. At present we have no explanation to offer for these variations.

Between 47.5 and 800 sq. A., the potential varies over the surface. The surface appears to be covered with two separate phases, one of which has the potential about 170 millivolts, the other an extremely small potential, about 4 millivolts. There is little doubt that these two potentials are those of the liquid expanded film in equilibrium with its surface vapour pressure, and the saturated vapour film. Adam and Jessop's measurements* indicated a region of constant surface pressure between 50 and 800 sq. A., in which these two surface phases co-exist in varying amounts. Fig. 4 gives some instances of the fluctuations observed near the greater of these areas; the abscissæ are the distances of the air electrode from the fixed barrier at the left-hand end of the film, which was put on at the areas stated from petroleum ether solution; the ordinates are the observed potential differences. Fluctuations are large at 710 and 405 sq. A., but at 940 and 1060 sq. A. they do not exceed 4 millivolts, which is not greater than the probable experimental error in these experiments, in which the surface is necessarily exposed to the air for a considerable time. It appears then that the region of varying surface potential

* 'Proc. Roy. Soc.,' A, vol. 110, p. 426 (1926).

ends in the neighbourhood of 800 sq. A., as would be expected from the surface pressure-area curves.

In the fluctuating region, the actual potential observed is largely a matter of chance; usually values as great as 150 millivolts could not be found at areas larger than 100 sq. A., although on one occasion a patch of about 150 millivolts was found at 300 sq. A., and transient high values were occasionally found between 100 and 300 sq. A. Above 300 sq. A. it appears that the size of the islands of coherent film are too small, in comparison with the portion of the surface whose potential is measured by the rather large air electrode, to give values as high as 150 millivolts. An attempt was made to estimate the proportions of the "gaseous" and coherent or "liquid expanded" film on the surface between 47 and 800 sq. A., as follows. The potential was measured at 27 equally spaced points on the surface; if the potential varied, owing to islands of film moving under the electrode during the measurement, the mean of the highest and lowest readings was taken. Assuming that the liquid phase had a potential of 170 millivolts, and the vapour phase one of 4 millivolts, then the fraction of the part of the surface under observation covered by liquid film should be $(\Delta V - 4)/(170 - 4)$, where ΔV is the mean potential observed. Taking this as constant within each of the 27 regions, and summing over the whole surface, a rough estimate of the fraction of the whole surface covered by liquid film could be made; this is given in Table I as Fr_{obs} . The theoretical

Table I.

Area per molecule. Sq. A.	Maximum variation in ΔV at any fixed point, mv.	Fr_{obs}	Fr_{calc}
70.6	128	0.43	0.64
123	117	0.24	0.34
214	99	0.18	0.17
293	72	0.072	0.10
405	66	0.045	0.06
710	21	0.02	0.01
940	Nil	Nil	Nil
1060	Nil	Nil	Nil
1380	Nil	Nil	Nil
1850	Nil	Nil	Nil

fraction (Fr_{calc}) was obtained from the formula $Fr_{calc} = 47(800 - A)/753A$, which assumes that the liquid phase has an area per molecule of 47 sq. A. and the vapour phase one of 800 sq. A., the latter figure being approximately the area at the right-hand end of the horizontal region in the surface pressure

curves of Adam and Jessop. The second column in the table shows the maximum fluctuation observed at any one of the 27 points during the 45 seconds or so required for an observation; these fluctuations disappear entirely above 800 sq. A. Fig. 5 shows the actual observations made at 70.6 sq. A. The agreement seems as good as could be expected, considering that the islands are moving about during the survey of the surface, and that the averaging of a fluctuating potential is a very rough proceeding.

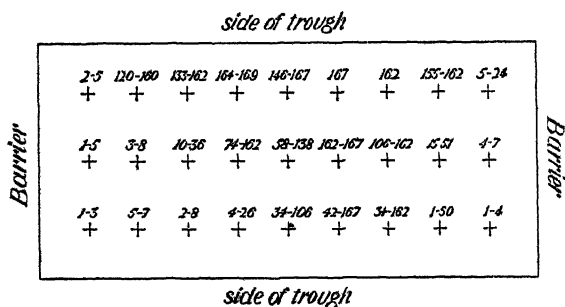


FIG. 5.—Myristic Acid on 0.01 N HCl. Temperature 12° C. Area per molecule : 70.6 square A.

The liquid expanded films have a slight tendency to stay up against any barrier with which they happen to be in contact; if a film is spread over the surface at an area of 47 sq. A., or less, and the barrier at one end moved away carefully after the film has fully spread, the film does not spread over the vacant space for some little time, but remains adhering to the barrier at the other end. This phenomenon is much more marked with the longer chain acids such as behenic, where the barrier may be moved away quite violently without moving the film.

Above 800 sq. A., up to 1850 sq. A., the potentials are too low to measure with accuracy. The average value over this region was about 4 millivolts and the results were not accurate enough to determine with certainty whether the potential falls off as the area increases, being subject to an error of at least 50 per cent.

Acids with Longer Chains.

The results with the acids of 16, 18 and 22 carbon atoms are shown in fig. 6. The films are all condensed. Fluctuations begin about 25 sq. A., where, according to earlier measurements of surface pressure, there are islands of condensed film in equilibrium with the vapour film. The lowest curves give

the surface pressure, the middle ones the values of ΔV . As soon as the condensed film is complete, and the area is further diminished, the potential ΔV increases slightly, but not as much as would be expected if the potential were directly proportional to the number of molecules in the film. The potential was uniform within 6 millivolts over the whole surface. The surface potential

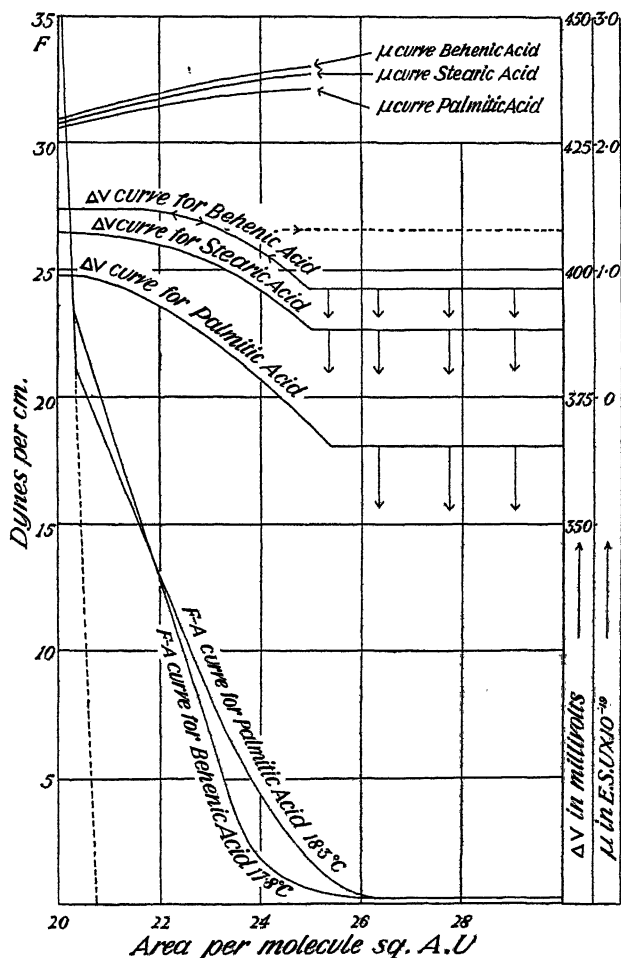


FIG. 6.

curve for palmitic acid could be obtained on compressing or decompressing the film; with stearic and behenic acids, however, on decompressing, fluctuations would often begin at an area of 23 sq. A. or less, the islands having the potential appropriate to that area on the ascending curve, corresponding to a film under a pressure of some 10 dynes per centimetre. This is shown as a dotted line on the curve for behenic acid. It would appear that the behenic acid, and to

a lesser extent the stearic acid film, can remain tightly packed after decompression, in an arrangement which can only be formed initially by compressing to about 10 dynes per centimetre.

In the fluctuating region, with behenic and stearic acids, it is much easier than with myristic acid to obtain definite potentials locally, corresponding to the pure coherent or pure vapour phases. This is no doubt due to the greater coherence of the longer chain acids in the films, making the islands larger and more permanent.

Calculation of the Dipole Moments.

Helmholtz showed that the change ΔV in surface potential, caused by a film containing n molecules per square centimetre, is given by

$$\Delta V = 4\pi n\mu; \quad (1)$$

μ is the average vertical component of the change in dipole moment of the part of the surface occupied by a film molecule, including all effects due to re-distribution of ions in the interior of the solution, and to re-orientation of water molecules.

The quantity μ , plotted at the top of figs. 3 and 6, varies surprisingly little for the various fatty acids, through the whole range of areas in which the expanded and condensed films occur. For myristic acid, it rises slightly from 2.15×10^{-19} e.s.u. per molecule at 47.5 sq. A. to 2.43×10^{-19} at 25 sq. A., a total variation of 12 per cent.; then falls off to 2.26 sq. A. at 22 sq. A., the film being practically condensed during this final falling off. The three longer chain acids (for which results are only available in the condensed region) have very slightly higher values of μ , which also show a slight falling off as the condensed film is compressed. Table II shows some results.

Table II.—Fatty Acids at Various Areas, on Dilute HCl.

—	No. of carbons.	Values of μ .			e.s.u. $\times 10^{-19}$.	
		20	22	25	34.7	47.5 sq. A.
Myristic.....	14	—	2.26	2.43	2.23	2.15
Palmitic	16	2.12	2.29	2.45	—	—
Stearic	18	2.16	2.36	2.57	—	—
Behenic.....	22	2.18	2.40	2.62	—	—

Our values for μ , for myristic acid, agree within 3 per cent. or less with the previously published values obtained by Guyot and by Schulman and Rideal; they are about 5 per cent. larger than those deduced from Frumkin's measurements.

In the vapour state, ΔV is too small to admit of accurate measurement; but μ appears to be within 30 per cent. of 1.4×10^{-19} e.s.u., *i.e.*, rather smaller than in the condensed and expanded states.

The question arises, is it possible to separate the total change in vertical dipole moment of the surface per molecule of the film into a term $\bar{\mu}$ expressing the vertical component of the dipole of each film molecule itself, and another term $\delta V/4\pi n$ due to redistribution of ions and to re-orientation of water molecules when the film is put on? We may write, with Schulman and Rideal,

$$\Delta V = 4\pi n\mu = 4\pi n\bar{\mu} + \delta V. \quad (1.1)$$

Differentiating (1.1), we have

$$\frac{d}{dn} (\Delta V) = 4\pi\bar{\mu} + 4\pi n \frac{d\bar{\mu}}{dn} + \frac{d}{dn} (\delta V), \quad (1.2)$$

and if we further assumed that the last two terms in (1.2) can be neglected, which would be the case if they were equal and opposite, or if $\bar{\mu}$ and δV did not change with n , we should have

$$\frac{d}{dn} (\Delta V) = 4\pi\bar{\mu}. \quad (1.3)$$

Schulman and Rideal, finding that the curve of ΔV plotted against n is practically a straight line over some regions of areas, have utilised equation (1.3) for calculating values of μ , calling these "absolute" values. We doubt very much, however, if this procedure is really a trustworthy method of eliminating the unknown quantity δV in equation (1.1). Our results show appreciable departures from a straight line, in the $\Delta V - n$ curve for myristic acid (dotted in on fig. 3, the units for the abscissæ, values of n , being at the top of the diagram), except between 47.5 and 37 sq. Å., and these departures become very great as the film becomes condensed. Table III shows the values

Table III.—Values of $\bar{\mu}$ and δV calculated by Equations (1.3) and (1.1) for Myristic Acid.

Area per mole- cule. Sq. Å.	n	$\frac{d}{dn} (\Delta V)$ (volts).	e.s.u.			δV (volts).
			$\bar{\mu}$	μ	$\mu - \bar{\mu}$	
47	2.13×10^{14}	8.95×10^{-16}	2.37×10^{-19}	2.16×10^{-19}	-0.21×10^{-19}	-0.0169
37	2.70×10^{14}	8.95×10^{-16}	2.37×10^{-19}	2.20×10^{-19}	-0.17×10^{-19}	-0.0173
30	3.33×10^{14}	11.1×10^{-16}	2.94×10^{-19}	2.3×10^{-19}	-0.64×10^{-19}	-0.0804
25	4.00×10^{14}	7.35×10^{-16}	1.95×10^{-19}	2.43×10^{-19}	$+0.48 \times 10^{-19}$	+0.0723
23	4.34×10^{14}	2.6×10^{-16}	0.69×10^{-19}	2.32×10^{-19}	$+1.63 \times 10^{-19}$	+0.267

of $\bar{\mu}$, calculated from equation (1.3), and corresponding values of δV , at various areas.

There is reasonable constancy in $\bar{\mu}$ and δV down to 37 sq. A., but very great variations below this area, δV changing sign at about 25 sq. A.; μ , however, varies but little over the whole range. In the case of behenic and stearic acids, at about 21 sq. A., $\bar{\mu}$ would be almost zero, according to this calculation, since ΔV scarcely alters with decreasing area per molecule.

It seems so improbable that the value of δV , the contribution to the surface potential change caused by re-distribution of ions and water molecules when the film is put on, would vary in this extraordinary way simply by crowding the molecules in the film, that we conclude this method of calculation of $\bar{\mu}$ and δV is not justified, *i.e.*, that it is not legitimate to neglect the last two terms in (1.2), over the whole range of areas from 47.5 to 21 sq. A.

Even apart from the above difficulties in particular cases, we are doubtful whether the use of (1.3) is justified. It is clearly not permissible to regard δV as independent of n at very large areas, for in the limiting case when the area is infinite, *i.e.*, there is no film on the surface, δV is necessarily zero. And even in the cases where ΔV varies linearly with n , it is not certain that the sum of the last two terms in (1.2) is zero. Constancy of $d/dn (\Delta V)$ could be obtained in other ways, *e.g.*, $d/dn (\delta V)$ might be constant or zero, but $\bar{\mu}$ might be varying in such a way that $\bar{\mu} + n (d\bar{\mu}/dn)$ was constant. Or, δV or $d/dn (\delta V)$ might be varying, $\bar{\mu}$ also varying so as to maintain the right-hand side of (1.2) constant. Such variations are by no means impossible; $\bar{\mu}$ will vary if the tilt of the dipole of the molecules is varied, as the molecules are crowded together. With Mr. Danielli and others we have obtained evidence of very large variations in tilt of the dipoles with derivatives of oestrin, and as the re-distribution of ions under the surface, and the orientation of the water molecules, which determine the value of δV , are controlled by the dipoles of the film molecules, it is very probable that alterations in δV will accompany changes in $\bar{\mu}$. Also, substituting the value of μ obtained from (1) and (1.1) in (1.2), it can easily be shown that the condition for the validity of (1.3) is that

$$\delta V = -4\pi n^2 (d\mu/dn), \quad (1.4)$$

i.e., that δV and $d\mu/dn$ have opposite signs.

In many cases, perhaps in most, the terms $\bar{\mu}$ and δV seem to be almost inextricably interconnected. Where there is a dissociable end group, which sends off an ion into the solution to a greater or less distance, are we to regard this ion as forming part of the dipole moment of the molecule itself, or as part

of the term δV ? Also, as will be shown in the next section of this paper, it seems likely that the total dipole moment of the undissociated carboxyl group can be altered, by rotation of the hydroxyl group about the line joining the oxygen of this group to the carbon, and that this rotation and change in total dipole moment is influenced by the distribution of ions underneath the film in the water. Hence it appears to us, in the light of the available data, difficult or impossible to separate the total vertical dipole moment μ into two terms, $\bar{\mu}$ and δV , one of which belongs to the film molecules and the other to the water molecules and the ions below. We think that the quantity μ is much more likely to be an indication of the orientation of the dipoles in the films, than the value of $\bar{\mu}$ given by (1.3) can ever be; it must always be remembered, however, that μ contains terms due to the underlying solution as well as to the film molecules alone. These terms are, however, probably controlled by the arrangement of the dipoles in the films.

Orientation of the Dipoles in the Surface.

The total dipole moment of the carboxyl group on an aliphatic chain is

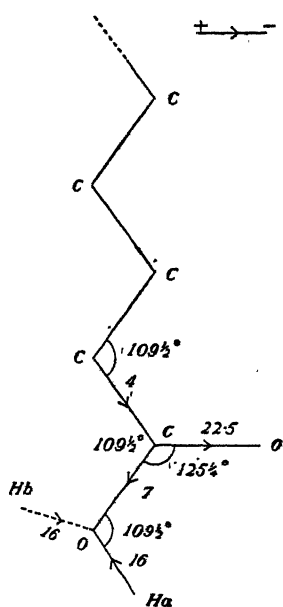


FIG. 7.—Direction of dipoles
+ —. Values in
e.s.u. $\times 10^{-19}$.

generally given as about 14×10^{-19} e.s.u.* It is not difficult, however, to account for the vertical component of about 2.3×10^{-19} which has been found by all workers on surface potential. Fig. 7 represents the lower atoms of the molecule, with the chain vertical. The carbonyl group might have a number of different positions, by rotation about the bond joining the α carbon to the carboxyl; we have chosen that in which the oxygen of the carbonyl is as high up as possible, which is that with this oxygen, and the plane of the double bond joining it to the carbon, in the plane of the surface. This fixes the positions of all the atoms of the end group except the hydrogen on the hydroxyl; this atom can rotate as the hydroxyl rotates about the line joining the carbon to the oxygen. The figure shows two extreme positions, in which the vertical component of the dipoles are a minimum and a

* Cf. Smyth and Rogers, 'J. Amer. Chem. Soc.', vol. 52, p. 1824 (1930).

maximum. With the hydrogen in the position marked *a*, the vertical component is -4.2×10^{-19} e.s.u.; in position *b*, it is $+13.2 \times 10^{-19}$, a complete change of sign. The positive sign is used to denote a vertical dipole in which the positive charge is uppermost. The values of Eucken and Meyer* for the various component dipoles, C—C, C=O, C—O, and O—H, as marked on the figure, were used, and the vectors added. The observed value for the vertical component, 2.3×10^{-19} , would be obtained if the hydrogen were rotated through an angle of about 75° from the position *a*. In the free carboxyl, we would expect the position *a*, owing to attraction between the negative oxygen and the hydrogen. On the hydrochloric acid solutions, according to Frumkin,† the chlorine ion is more adsorbed than the hydrogen, rendering the surface more electronegative than the interior; this might attract the hydrogen towards position *b*, which is rather higher up in the position we have chosen, with the chain vertical.

Clearly the vertical component deduced from the measurement of surface potential is consistent with the values for dipole moments obtained in other ways. Not much stress can, however, be laid on the agreement, since such a wide range of values is obtainable by rotation of the hydroxyl; and also the re-distribution of ions and re-orientation of the water molecules by the presence of the film has not been taken into account.

The small change in vertical component of the dipole moment per molecule, as the film is compressed from the liquid expanded film at practically no surface pressure, to the condensed film at high pressure, indicates an approximately constant orientation of the end group to the surface throughout this region. One of us has suggested‡ that, in the liquid expanded films, the lateral adhesion between the chains characteristic of the condensed film has broken down, the chains oscillating and whipping about while the end groups adhere sufficiently to maintain the cohesion of the film. Such oscillation is possible, through rotation of the atoms round the carbon to carbon linkages; at low temperatures, in the condensed films, the thermal agitation is insufficient to overcome the tendency of the chains to lie side by side; in the expanded films, the greater part of the adhesion between chains has broken down, but that between the heads remains. The approximate constancy of orientation of the dipoles associated with the carboxyl groups throughout the region where the oscillations occur is additional evidence in favour of this theory.

* 'Phys. Z.,' vol. 30, p. 397 (1929).

† 'Z. phys. Chem.,' vol. 109, p. 34 (1924).

‡ "The Physics and Chemistry of Surfaces," p. 75 (1930).

If the observed fall in μ by some 40 per cent., on passing to the vapour state of the films, is genuine, it indicates some re-orientation of the carboxyl groups as the chains lie flat (the flat position is the accepted orientation of the whole molecule in the gaseous films). A model of the molecule shows that any desired orientation of the carboxyl group could easily be obtained, by rotation of the whole carboxyl group, or of the hydroxyl, with a flat position of the chain. The values of μ in the gaseous state are, however, very rough.

We are much indebted to Professor F. G. Donnan, F.R.S., for his constant interest in this research; to Professor E. K. Rideal, F.R.S., and Drs. Schulman and Hughes, for cordial interchange of ideas during the progress of the work; and to the Government Grant Committee of the Royal Society for a grant which defrayed part of the expense of the work. One of us (J. B. H.) is indebted to the Department of Scientific and Industrial Research for a maintenance grant.

Summary.

An apparatus for simultaneous determination of the change in contact potential, ΔV , at the interface between aqueous solutions and air, caused by a surface film, and the surface pressure of the film, is described.

With the fatty acids from 14 to 22 carbon atoms in length, ΔV appears to increase very slightly with increasing length of chain. There is a steady rise in ΔV , nearly, but not quite, proportional to the increasing number of molecules in the film, as the liquid expanded film is compressed. The rise in ΔV on compressing the condensed films is less than that corresponding to the number of molecules in the film. The total vertical component of the change in dipole moment of the surface, per molecule of the film, rises slightly as the expanded film is compressed, and falls as the condensed film is compressed; the variations do not, however, exceed 15 per cent. This indicates that the orientation of the carboxyl groups in the film does not change much on compressing through the whole of the regions of liquid expanded and condensed film.

In the region of constant surface pressure, where earlier work had indicated the probable co-existence of liquid expanded and vapour film, exploration of the surface indicates that these two phases are the only ones present, and that they co-exist in approximately the theoretical proportion.

Rough measurements indicate that the vertical component of the dipole moment of the carboxyl group in the surface is somewhat diminished in the vapour state, indicating some re-orientation of the end group.

It is shown that the observed vertical components of the dipole moment are compatible with existing data on the various component dipoles of the carboxyl group.

On the Surface Potentials of Unimolecular Films. Part IV.—The Effect of the Underlying Solution and Transition Phenomena in the Film.

By J. H. SCHULMAN and A. H. HUGHES, Department of Colloid Science, Cambridge.

(Communicated by E. K. Rideal, F.R.S.—Received June 14, 1932.)

INTRODUCTION.

In previous papers of this series (1) it was shown that the method of surface potentials could be developed to give quantitative measurements of the physico-chemical properties of unimolecular films at an air liquid interface.

Several problems of fundamental importance arise when a closer investigation is made of the true meaning of the phrase "surface potential of a unimolecular film."

In this work a study has been made of the phenomena attached to the deposition of a film on aqueous solutions, with special reference to the mutual influence of the solution and the polar head-group of the film-forming material.

This was considered necessary as a prelude to the general application of the method of surface potentials to the study of surface reactions and systems of biological importance. Work of this nature is in progress, and it is hoped will form the subject of a future communication.

It had already been noticed in Part III for the fatty acids that the graph of the surface potential (ΔV) against the number of molecules per square centimetre (n) showed a parallel displacement on alteration of the pH of the underlying solution over certain ranges where the state of the film remained constant. The relation of this movement to the ionisation of the head groups of the film was discussed, but it was now thought advisable to carry out the

experiments over much larger ranges of p_H , and to extend them to polar head groups other than the carboxyl.

EXPERIMENTAL.

The apparatus used was essentially the same as that already described in Part I,* with the addition of an arm moving on a horizontal plate holding the polonium-coated air electrode, so constructed that it moved at the required constant height over the whole of the liquid surface. This modification was found essential when investigating the homogeneity of the film.

The weak link in the accuracy of this type of work is the estimation of the area per molecule occupied by the film-forming substance. The procedure hitherto employed involves the use of a small pipette to place on the surface two or three drops of a solution of a known weight of the substance in a known weight of petrol ether. The weight of a large number, say 50, of the drops is measured; hence the weight of one drop and the weight of the substance on the surface. This method is liable to grave errors, due to the rapid evaporation of the petrol ether, and to the assumption of the uniformity of the drop size. The error in using only one drop of a calibrated solution may be as much as 10 per cent. off the mean value of 50 drops. In the earlier work no allowance was made for this error.

A new instrument was therefore devised to give molecular areas reproducible to 1 per cent. It is a micro-pipette consisting of a long fine capillary tube, calibrated for volume by a mercury thread, and fixed in a protecting glass tube. The petrol ether solution is drawn up to a definite mark on a scale placed along the capillary, by rotation of a piston fitting tightly into a rubber gland in the widened end of the capillary tube. The requisite approximate number of drops is then expelled on to the surface, but the *volume* of solution lost can be found accurately from the calibration of the capillary. The actual size of each drop is thus immaterial.

Dr. R. J. Fosbinder, of the Cancer Research Institute, Philadelphia, very kindly pointed out to us that there was an instrument on the market which could be adapted for accurate measurements of these small volumes. This is the Agla Micrometer Syringe (Burroughs, Wellcome & Co., Ltd.), which was recalibrated, and found to give consistent results to 1 per cent. on the area per molecule.

It is relevant to observe here that three different types of apparatus have been used, and they have given results for the surface potentials identical to within

* Schulman and Rideal, 'Proc. Roy. Soc.,' A, vol. 130, p. 259 (1931).

1 mv., while the results are also unaltered by the use of many different ionisation electrodes of various shapes and strengths, together with use of either silver, silver chloride or calomel reversible electrodes for the underlying solutions.

MYRISTIC ACID AND DODECYL-ALCOHOL. THE EXPANDED STATE.

It was shown in Part III* that over the region of ionisation p_H 4– p_H 6.5, the films of palmitic acid existed in the liquid condensed state, and that the value of the electric moment per molecule ($\bar{\mu}$) calculated from the slope of the curve remained constant for that state in spite of a considerable change in the value of ΔV at any given area per molecule.

On further increase in ionisation the state and slope of the curve changed, but $\bar{\mu}$ again remained constant over a range p_H 6.5– p_H 8 in the solid condensed state, the values in the two states being $1.56 \cdot 10^{-19}$ e.s.u. and $3.2 \cdot 10^{-19}$ e.s.u. respectively.

As will be seen from figs. 1 and 2, the same general behaviour is shown by myristic acid and by dodecyl alcohol for the expanded state. Myristic acid gives a series of parallel lines from N HCl to distilled water, with $\bar{\mu} \doteq 2.5 \cdot 10^{-19}$ e.s.u., while for dodecyl alcohol the expanded state continues to the most alkaline solutions studied (N.NaOH). In the case of dodecyl alcohol, the values of “ n ” and hence of the slope of the curve are untrustworthy, especially on alkaline solutions, owing to the solubility of the film. Thus the films can be compressed to apparent areas per molecule much less than the limiting area of 18.5 sq. A. per molecule for vertically oriented molecules. The more quickly the readings are taken after the deposition of the film, the nearer to 18.5 sq. A. is the observed area at which the film collapses.

The value of $\bar{\mu}$ in the expanded state for dodecyl alcohol is approximately $2.0 \cdot 10^{-19}$ e.s.u.

The next point of interest to be observed from the figures is the general contraction of the film as the alkalinity is increased. Thus the limiting area of the transition of the expanded state to the vapour state for myristic acid, fig. 1, “ α ,” changes from 55.0 sq. A. on 4 N HCl to 33.0 sq. A. on water, although on N/10 and N/100 HCl the area is the same, viz., 45.0 sq. A., the curves on these two solutions being identical. For the alcohol on 4 N HCl the expansion is so marked that a “vapour expanded” film (curve A) appears to be formed, the film remaining homogeneous up to at least 400 sq. A. per

* Schulman and Rideal, *loc. cit.*

molecule at about 20 mv. On the alkaline side, however, though the acid shows a transition to a solid condensed film at p_H values greater than 7, and ultimately forms soap films where the value of ΔV is reversed in sign, the alcohol films show no change from N/100 HCl till a pH of about 10 or 11, being still in the expanded state.

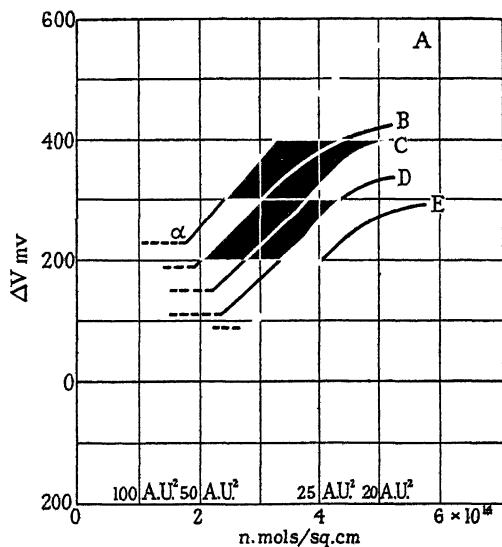


FIG. 1.—Myristic Acid, 15° C.

- A.—4 N HCl.
- B.—N H Cl.
- C.— 10^{-2} and 10^{-1} N HCl.
- D.— 10^{-4} N HCl.
- E.—Distilled water.

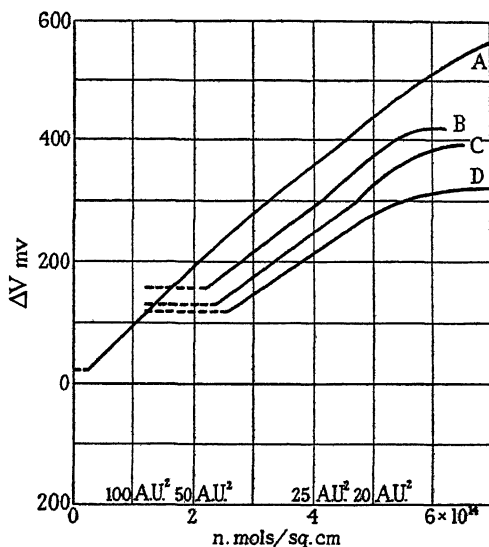


FIG. 2.—Dodecyl Alcohol.

- A.—4 N HCl.
- B.—N HCl.
- C.—0.01 N HCl and 0.001 N NaOH.
- D.—0.1 N NaOH.

With reference to the vapour expanded films a curve is included for comparison, showing the behaviour of ethyl myristate on N/100 HCl, fig. 3. Here the film remains homogeneous down to the lowest measurable surface potentials, at areas larger than 500 \AA^2 per mol, in direct contrast to the "expanded" films which break down to inhomogeneous two-phase systems.

Vapour expanded films, as shown by N. K. Adam, are expanded films above the critical temperature.

TETRADECYL ALCOHOL.

Tetradecyl alcohol at 32° C. forms an expanded film which shows an essentially similar behaviour to that of dodecyl alcohol at room temperatures, but accurate results are again not possible owing to the solubility of the film.

It is interesting to note, however, that tetradecyl alcohol, fig. 3, which gives liquid condensed films at room temperatures from N HCl to N NaOH, gives on 4N HCl an expanded film with a limiting area at 55 sq. A. per molecule, again indicating the very marked effect of concentrated hydrochloric acid in expanding the film, as was observed on dodecyl alcohol.

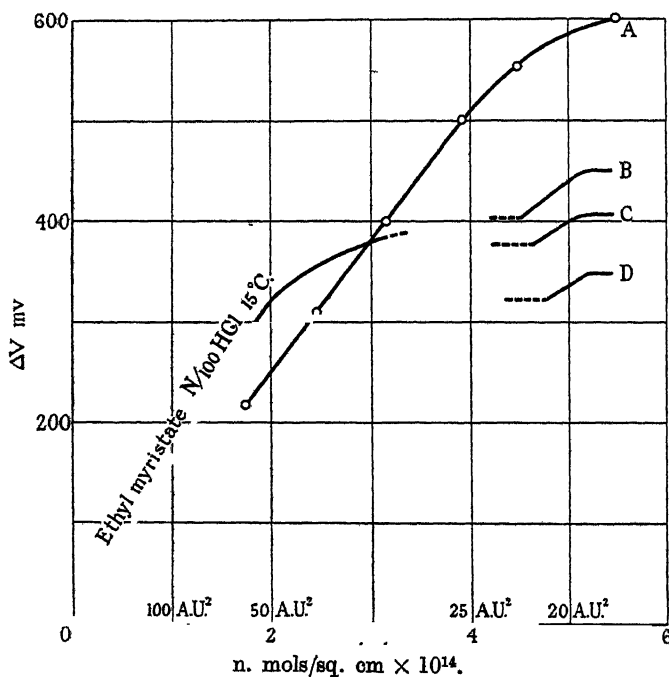


FIG. 3.

The effect of p_H on films of the alcohols is again reflected in the condensed films of tetradecyl alcohol, which show the same behaviour from N/100 HCl to p_H 10 or 11, but contract both in area and potential at higher p_H values. The slope of these lines is constant, and gives $\bar{\mu} = 1.1 \cdot 10^{-19}$ e.s.u. The condensed films of alcohols will be discussed in detail later.

For the liquid condensed films of palmitic acid a similar expansion has been noticed on increasing the acidity of the underlying solution (Part III) (*loc. cit.*), as also observed in the force/area curves by Lyons and Rideal.*

Beyond the superior limiting areas of the expanded state the films become inhomogeneous, the value of ΔV showing large fluctuations over the surface. Below the limiting area the linear character of the $\Delta V/n$ curves is preserved down to areas of about 35–30 sq. A. per molecule, but at this point transition

* 'Proc. Roy. Soc.,' A, vol. 124, p. 322 (1929).

to the liquid condensed state begins. From 35–25 sq. A. per mol down to about 25 sq. A., the curve does not seem to be definitely reproducible, and fluctuations in potential can sometimes be detected over the surface, of about 20–30 mv., though this is dependent on the rate of compression of the film. By about 25 sq. A., however, the slope of the curve has decreased markedly and is much smaller in the liquid condensed state than in the expanded.

It is important further to notice that the values of the intercept (ΔV_u) of the $\Delta V/n$ curves on the axis of ΔV obtained through extrapolation of the straight portion of the curve can be placed broadly in two classes:—(1) Expanded or Vapour-Expanded Films, ΔV_u negative, and (2) Liquid Condensed Films, ΔV_u positive.

COMPARISONS OF ACIDS, ALCOHOLS AND ETHERS.

In order to cast more light on the mutual influence of the polar head-group and the underlying solution, a long-chain ether was examined over a large range of p_H . Octadecyl methyl ether was chosen, this compound forming liquid condensed films on all solutions so far examined at room temperatures.

The results (a) are compared with those for the alcohols (b) and the acids (c) in fig. 4, in which the p_H of the underlying solution is plotted against ΔV_u , the limiting surface potential obtained by placing an excess of the substance on the surface.

The sharp rise of potential on concentrated hydrochloric acid is seen to be common to all three substances, but whereas the ether, definitely non-acidic in character, acquires almost identical values for the surface potential at p_H 2 as at p_H 12 (453 mv. and 448 mv. respectively), the alcohol shows a marked falling off at p_H 12 and at greater p_H values.

The behaviour of the acid is very definite, ionisation of the film setting in between p_H 3 and 4, with a complete reversal of sign beyond p_H 9.

It was noticed in the case of the alcohols on alkaline solutions that the initial value of ΔV was always some 30–40 mv. higher than the final equilibrium value, the fall taking place in about three minutes. It is possible that this represents the actual course of ionisation or partial ionisation of the OH group, the value of ca. 400 mv. from p_H 1–10 being that of the unionised alcoholic group. This effect is similar to that occurring during the ionisation of the acids.

INFLUENCE OF SALTS.

A further series of experiments was carried out to investigate the effect on the surface potential of the film of solutions containing a variety of ions.

Salts were used which both raise or lower the surface tension of water. Thus, for tetradecyl alcohol KCl had no effect at any concentration from $N/100$ to 3.5 N , but LiCNS, $N/2$, caused an increase of 25 mv.; potassium ferricyanide, sodium citrate, ethylamine hydrochloride, an increase of ca. 10 mv.; and

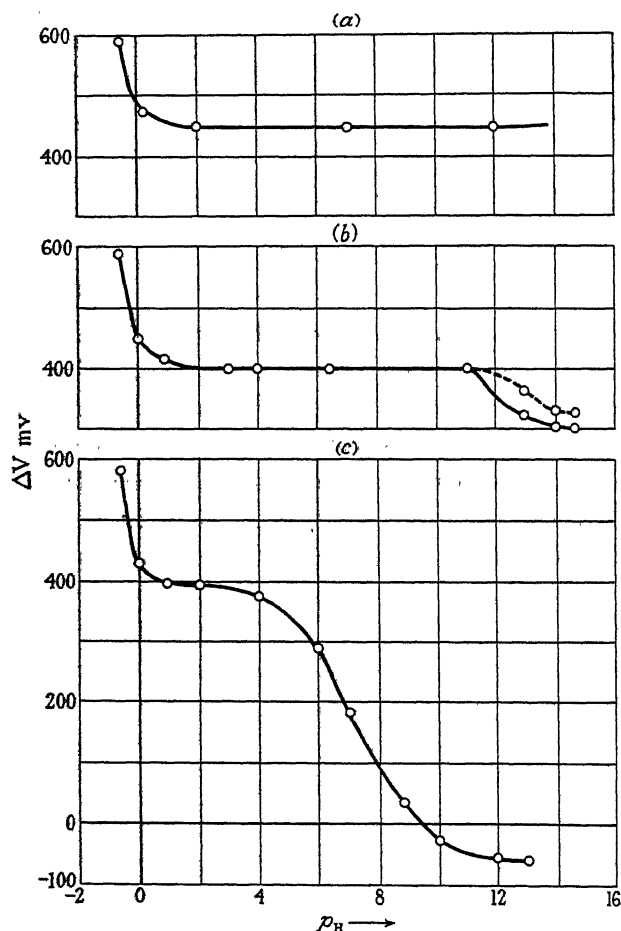


FIG. 4.

$\Delta V_{\infty} / p_H$ curve 15°C .

ΔV_{∞} corresponds to an area approximately 20 A^2 per mol.

a. Octadecyl methyl ether; b. tetradecyl alcohol; c. Myristic acid.

$\text{Al}_2(\text{SO}_4)_3 - N/100$ p_H 4.7 — an increase of ca. 35 mv., all reckoned from the value of 400 mv. on distilled water. In no case has a decrease been observed, except where it is reasonable to assume ionisation of the film. For the ether no alteration of the potential could be detected due to the nature of the ions

in the solution, the value of ΔV_{∞} remaining constant at 448 mv. on all these solutions.

DISCUSSION OF RESULTS.

In summarising the experiments on the expanded and condensed films of the acids and alcohols, with reference to the effect of the underlying solution, the following conclusions may be drawn : (1) In the expanded state for the acids and the alcohols there is little alteration in the slope of the $\Delta V/n$ lines over a very large range of p_H of the underlying solution, although the actual value of "n" may vary to the extent of some 250 mv. The extrapolation of these lines to the axis of ΔV give negative values of ΔV_u . (2) In the condensed film of acids and alcohols, parallel straight lines are again obtained over large p_H ranges, the ΔV_u values here being positive.

We may apply the Helmholtz equation and write

$$\Delta V = 4 \pi n \mu \tag{1}$$

where μ is the total vertical component of the polypole system produced at the air liquid interface by the presence of one molecule in the film.

Where the $\Delta V/n$ graph is linear we may further write

$$\Delta V = an + b \tag{2}$$

"a" and "b" being constants.

Combining equations (1) and (2), we obtain

$$4 \pi n \mu = an + b$$

or

$$\mu = \frac{1}{4\pi} \left(a + \frac{b}{n} \right) \tag{3}$$

i.e., μ can be regarded as composed of two portions :—

(i) $a/4\pi$ independent of "n" and independent of the p_H of the solution in any one state of the film (the $\bar{\mu}$ of previous papers).

(ii) $b/4\pi n$ varying inversely as "n" and dependent on the p_H of the solution ($b = \Delta V_u$ of previous papers).

Thus (α) if "b" is positive, then, as "n" increases, μ decreases (condensed state); but

(β) if "b" is negative, as "n" increases μ increases (expanded state).

Thus the simplest interpretation of this analysis is that in any one state of the film the total μ consists of a portion $\bar{\mu}$ or $a/4\pi$, which is that of the film-forming material itself, and is constant for that state; and of a portion

$b/4\pi n$, which represents the manner in which the contribution due to the underlying solution is affected by compression of the film. As has already been noticed, the order and sign of the value of ΔV_u are characteristic of the state of the film, but too great stress must not be placed on the numerical values of ΔV_u , since these are obtained by extrapolation of the $\Delta V/n$ curves.

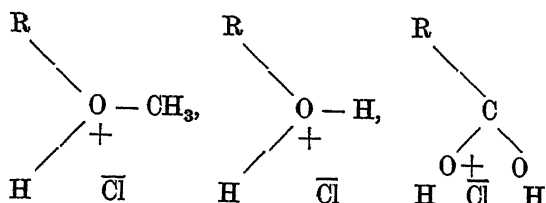
It is clear, however, that in compression of the expanded state the contribution of the underlying solution is *increasing* in the manner given by the term $b/4\pi n$ when " b " is negative.

Whether we may consider " b " as being due to an orientation of solvent molecules in the neighbourhood of the polar head, together with a resultant dipole moment, or due to preferential adsorption of the ions in the Gibbs' layer, is at present difficult to decide; the latter effect is definitely present, as shown by the effect of different salt solutions on the film potentials.

Further, on compression of the condensed state the contribution of the underlying solution is *decreasing*, since " b " is now positive. This may be ascribed either to an expulsion of oriented solvent molecules from the vicinity of the head groups when these become close packed in the condensed state, or to a marked change in the distribution of the ions in the Gibbs' layer.

Referring once more to figs. 1, 2 and 3, in regard to the effect of hydrochloric acid solutions on the film, an increase of the acid concentration above N/10 results for the acid, alcohol or ether in a gradual expansion of the range of homogeneity of the films, together with a general rise of potential. The change of $\bar{\mu}$ is, however, not large, increasing from $2.5 \cdot 10^{-19}$ e.s.u. for myristic acid in the expanded state on N/10 or N/100 HCl to $3.0 \cdot 10^{-19}$ e.s.u. on 4 N HCl; and from $1.9 \cdot 10^{-19}$ e.s.u. on N/10 or N/100 HCl to $2.1 \cdot 10^{-19}$ e.s.u. on 4 N HCl for dodecyl alcohol, also in the expanded state. For the ether the range of existence of the condensed film is so small—about 21 sq. A. to 19 sq. A.—that the slopes cannot be estimated with accuracy.

This general behaviour on solutions of hydrochloric acid may be interpreted as being due to the formation of an ionisation complex on the polar head-group. In the extreme case one might postulate the formation of oxonium compounds of the type:—



for the ether, alcohol and acid respectively. This results in an increase of the dipole moment of the film if carried to completion, and a definite alteration both in moment and in distribution of the ions in the Gibbs' layer if the formation of this type of complex is only partial. These effects are observed. The greater the concentration of HCl in the underlying solution, the greater the change in moment. It must be emphasised that the values here obtained for the dipole moment refer only to the *vertical component* of the polypole system. It is impossible to specify the total change which may occur in the head group and its surrounding ion atmosphere.

For the ether, alcohol and the acid head group, the dissociation of the HCl complex appears to be complete when the HCl concentration in the solution has sunk to $N/10$. Thence onwards, with increasing p_H , the ether shows no further change in the head group or the underlying solution. For the alcohol no change is observed up to a p_H of about 10, but beyond this point ionisation sets in, and can be followed by the sharp falling off in the film potentials; and for the acid this falling off occurs at a much smaller p_H value, setting in at about p_H 4.

The general form of ionisation curve is very similar to that found in an electrometric titration in the bulk phase, fig. 4.

The behaviour of the alcohol indicates that it is acting as an acid with a dissociation constant of about 10^{-13} . It is interesting to compare this result with those obtained by Michaelis and Rona* for the acid dissociation constants of sugars, which were found to give $K_a = 10^{-13}$.

It would be of great interest in this connection to examine the cataphoresis of these acids and alcohols over a large range of p_H , and also in the presence of different types of ions.

The work of G. Köhler† on the electric endosmose of liquid through a palmitic acid diaphragm has shown that there is no relative movement at p_H 3.6, but at greater p_H values ionisation and movement takes place. No work appears to have been carried out on the alcohols, or on the acids on more strongly acid solutions.

The results we have obtained on the general effect on the surface potential of the film due to the presence of various salts in the underlying solution, show that small changes in the values of ΔV can be effected, causing a small parallel displacement to the $\Delta V/n$ curve. For tetradecyl alcohol the changes were in no case greater than about 35 mv. (on $N/100$ aluminium sulphate), and

* 'Biochem. Z.,' vol. 49, p. 232 (1913).

† 'Z. phys. Chem.,' A, vol. 157, p. 134 (1931).

always in the same sense, namely, an increase in the original potential obtained on distilled water.

The sign of the potentials of these films shows that the dipole moment is positive above and negative below. It is reasonable to suppose that adsorption of ions in an oriented layer under the film will take place by electrostatic adsorption, giving a second layer positive above and negative below. This results in an increase of the observed values of ΔV .

These small variations in potential are comparable in magnitude to the values obtained for zeta potentials in cataphoretic experiments, as distinct from the order of magnitude of the changes occurring when ionisation of the film is taking place, which may amount to 200 or 300 mv.

It is clear that if there is no orientation of dissolved substances in the underlying layer, these will have no effect on the potential of the film, even though they may produce large changes when present alone at the air-liquid interface. This is exemplified by the behaviour of a long-chain methyl ether, the film potential of which is independent of the presence of a large number of substances in the underlying solution.

These considerations permit us to draw a distinction between oriented adsorption in the Gibbs' layer and actual chemical changes occurring in the film.

TRANSITIONS.

The Vapour/Expanded Transition.

A moving electrode was employed in investigating the limiting areas of the vapour/expanded transition. It was found that the films at areas beyond 45 A^2 for myristic acid on $\text{N}/100 \text{ HCl}$ were inhomogeneous. This fact was first pointed out to the authors by Adam and Harding, who, in the course of experiments on surface potentials, had found that by blowing on the surface very large fluctuations occurred in the potentials over this range. As an apparatus had already been constructed with a movable electrode for investigating the limiting area of the condensed states, where large fluctuations had already been noticed (Part III), the observation of Dr. Adam was soon confirmed.

In Part I the polonium source was fixed stationary over that end of the trough towards which the film was compressed, and inhomogeneity was only noticed at areas greater than about 100 A^2 per mol. This was found to be due to the expanded form of the myristic acid existing in large islands, which on compression collected under the polonium electrode long before the surface was

homogeneous. These islands had a maximum potential of 152 mv., and the vapour film existing between them possessed a very low surface potential of approximately 2 mv.

Since the surface pressure remains constant over this region and the potentials of the two forms of the film present are so very different, the limiting area of the phase with the larger surface potential was easy to determine with accuracy.

The limiting area of the vapour film, on account of its very small surface potential, was much more difficult to determine. The method employed was to observe the area at which the islands of the expanded film disappeared and the whole surface became a homogeneous vapour film. The polonium was moved systematically over the surface, and all non-uniformities with their potentials were localised. These islands were then closely observed over a contraction and expansion of the film.

It appeared at first that the surface became uniform with a potential of about 2 mv. at areas of 300 \AA^2 , and the curves seemed consistent with those already published (Part I). But on closer observation it was found that there was considerable non-uniformity of about 10–20 mv. within 1–2 cm. of the edges of the trough. It may be added that the uniformity of the potential of the liquid surface was examined in the absence of the film and found to be uniform within 2 mv. even up to a few millimetres from the edges of the trough.

That the liquid portion of the film in the transition region should tend to collect along the edges of the trough was apparently due to the spreading of the petrol ether solution. It was noticed in a blind experiment using pure petrol ether alone, that any surface contamination was swept to the edges of the trough by the petrol ether. This was a convenient and sensitive method of checking the purity of the surface.

Non-uniformities were quite definitely present on the surface from areas of 700 \AA^2 per mol. to 45.5 \AA^2 at 15°C . At areas of 1000 \AA^2 per mol. changes of only 1–2 mv. from the zero value were noticed over the whole surface. Assuming that the surface is uniform at 45 \AA^2 per mol. at a potential of 152 mv., one may calculate the fraction of the total surface at any larger mean area per molecule which should be covered with the 152 mv. species of film, if the latter is unchanged on expansion. This may be compared experimentally with the area so covered, as found by moving the polonium electrode over the surface. The values so observed were always well below the calculated value for areas beyond 100 \AA^2 .

The explanation of this is probably that the islands of the 152 mv. expanded film are becoming too small to occupy the area subtended by the whole solid angle of the ionisation region of the polonium source.

In summarising, one may say that the range of inhomogeneity agrees very well with the limits of the transition already found by Adam and Jessop* for the myristic acid vapour/expanded transition, namely, from 850 \AA^2 — 48 \AA^2 per mol. at 16° C . This inhomogeneity of the transition shows there is no necessity for assuming the pre-Homalic state (*cf.* Part I).

The low values of ΔV for the vapour state at 850 \AA^2 may be compared with the values for the vapour films of soluble substances. Thus, butyric acid gives a ΔV of 22.5 mv. when “ n ” deduced from the surface tension depression is $0.35 \cdot 10^{-14}$ mols./sq. cm., corresponding to an area per molecule of about 300 \AA^2 per mol. The vapour expanded films of dodecyl alcohol on concentrated HCl and of ethyl myristate on N/100 HCl give the same order of ΔV (20 mv.) at these large areas.

The moving air electrode permits of an interesting experiment in that one can follow the surface solution of a crystal of, say, myristic acid placed in the liquid surface.

In Part I, curves were shown of the change of surface potential with time at a fixed point on the surface during the process of surface solution. When a large island of the expanded state (with a potential of 152 mv.) arrived under the polonium electrode the surface potential remains constant until the end of the transition, when the whole of the surface is covered with this form. Then the pressure increases and the potential rises proportionately until the equilibrium spreading pressure is reached.

If the polonium electrode is placed next to the crystal immediately it has been placed in the surface, violent fluctuations of potential (0–100 mv.) take place within 1–2 cm. of the crystal. Beyond this point there is a region of uniform potential at 152 mv. in the form of a gradually expanding wave with a sharp boundary. In front of this there is a potential of only 1–2 mv., *i.e.*, the vapour film.

The rate of spreading of the 152 mv. form (the uncompressed expanded film) can be readily followed with the moving electrode. This method of following the spreading of films could be made of general application.

The fluctuations of potential in the immediate vicinity of the crystal were explained by observations with a dark-field ultramicroscope of the type

* ‘Proc. Roy. Soc.’ A, vol. 110, p. 423 (1926).

designed by Zocher and Stiebel.* Here concentric rings of minute crystallites, similar to those appearing in the "point structure," at film collapse, were seen to spread off from the surface of the crystal and dissolve away within a few millimetres.

The Expanded/Condensed Transition.

The moving electrode was again of use in observing the transition of the film from the expanded into the condensed state. On compression of a film of myristic acid on N/100 HCl to areas smaller than about 35 A^2 per mol., the potentials begin to show fluctuations over the surface to the extent of some 20 mv.

The film becomes uniform again at smaller areas, below about 27 A^2 per mol., and on expansion the process is reversible.

These fluctuations do not always occur. Thus, on depositing the film on the surface directly at an area of about 30 A^2 per mol., the film is uniform and the potential follows the repeatable course shown in fig. 1. The fluctuations appear, however, when the film is compressed from larger areas. Whether the non-uniformity really indicates the presence of two distinct phases, liquid expanded and liquid condensed, or whether it is due to local collapse of the film, is difficult to decide. From the measurements of surface pressures (N. K. Adam†), a definite break in the Force/Area curves is always found at areas of about 35 A^2 per mol. at 15° C .

At smaller areas than about 30 A^2 a gradual change in the slope of the $\Delta V/n$ curve occurs, and the value of $\bar{\mu}$ decreases, the final portion of the curve corresponding closely with that of the liquid condensed films of palmitic acid under the same conditions.

It is worthy of note that myristic acid at 5° C . gives a curve at slightly higher values for ΔV following the upper portion of that obtained at room temperature, the limiting area now being 28 A^2 ; the change of slope ($\bar{\mu}$) is also noticed as the film becomes condensed. There is thus a marked change in the character of the film, as shown by the surface potentials as well as by the surface pressures.

THE VAPOUR/CONDENSED TRANSITION.

Transitions in the Condensed State.

In Part III, and in the present work, reference has been made to the condensed films of palmitic acid and of tetradecyl alcohol. Further experiments

* 'Z. phys. Chem.,' A, vol. 147, p. 401 (1930).

† "The Physics and Chemistry of Surfaces," chap. 2, London (1930.)

on this type of film have been made with hexadecyl and eicosyl alcohols to examine the influence of the nature of the polar head and of the chain-length on the general behaviour in the condensed state.

The range of existence of the condensed state is relatively small—at the most from about 23 \AA^2 to 18.5 \AA^2 per molecule in these cases—and the limiting factor in the accuracy of these experiments is the estimation of the molecular areas. These can be measured to 1 per cent.

The Force/Area curves for hexadecyl and eicosyl alcohols were determined for comparison of the limiting areas of the condensed state obtained in this way with those obtained by the method of surface potentials.

FORCE/AREA CURVES.

Fig. 5 shows the behaviour of hexadecyl alcohol on N/100 HCl at room temperature, and is typical of the condensed state, the graph being clearly

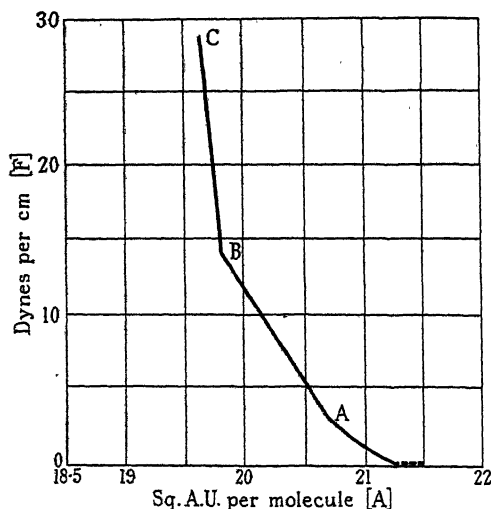


FIG. 5.—Force/Area Curve. Hexadecyl alcohol N/100 HCl 15°C .

divided into two straight lines, AB and BC, representing a high and a low compressibility respectively. The film is liquid even up to the highest compressions, a point of some interest to which reference will be made later.

Fig. 6 gives the Force/Area curves of eicosyl alcohol also on N/100 HCl. At low temperatures this alcohol forms a definitely solid film, and its behaviour was, therefore, examined over a range of temperatures from 4°C . to 31°C ., with a view to correlating the phase change in the film, as observed by the

motion of talc particles placed on the surface, with any change in the form of the force/area curves.

At 4° C. the film is solid, and a noticeable surface pressure is first obtained at an area of 21.0 A² per mol., and a linear increase in pressure is obtained down to 19.1 A² per mol. in the AB form of high compressibility. The form of low compressibility and high pressures is unstable at this temperature, and the film readily collapses above 10 dynes per cm.

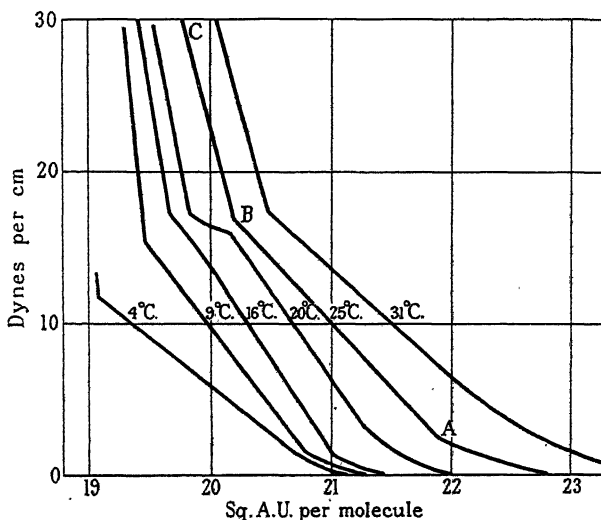


FIG. 6.—Force/Area Curves. Eicosyl alcohol [C₂₀] N/100 HCl.

Up to 16° C. the same type of curve is obtained as on hexadecyl alcohol, and the second portion, BC, can be followed from the change point B, at 17 dynes per cm. to pressures above 30 dynes per cm. The AB film is now liquid and the BC film solid.

In the neighbourhood of 20° C. the force/area curves show a peculiar arrest level from 20.1–19.8 A² per mol. just before the steep rise in pressure along BC. Above 22° C. the film is liquid throughout, and gives the usual type of curve. There is no sharp melting point of the film. A gradual increase of the molecular areas with rising temperatures is obtained of about 10 per cent. for a range of 27 degrees. The results of Adam and Dyer* on the alcohols appear to agree with the results here given at room temperature.

SURFACE POTENTIAL CURVES.

The surface potentials (ΔV) in the condensed films show a similar change of form, at a definite area per molecule, as the force/area curves, and provide more

* 'Proc. Roy. Soc.,' A, vol. 106, p. 694 (1924).

detailed information as to the nature of the film. In the case of hexadecyl alcohol, the effect of the underlying solution, and in the case of eicosyl alcohol the effect of alteration of temperature, has been investigated. The values of ΔV are plotted as usual against " n ," the number of molecules per sq. cm., instead of " A ," the area per molecule, since the slope of the $\Delta V/n$ curve gives information as to the dipole moment of the molecule. The results on hexadecyl alcohol are given in fig. 7.

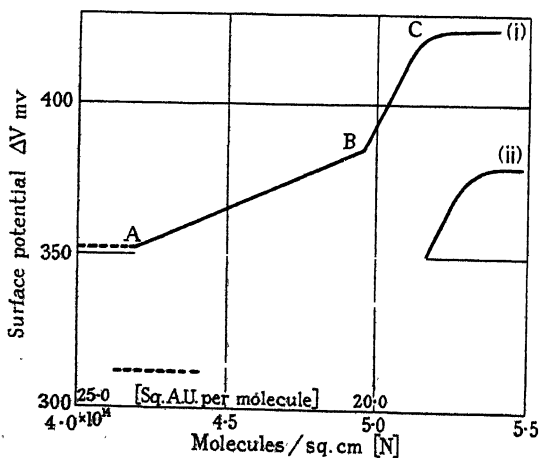


Fig. 7.—Hexadecyl alcohol, 16° C. (i) N/100 HCl; (ii) N/10 NaOH.

At areas larger than about 23 \AA^2 per mol. the film is inhomogeneous, the potentials fluctuating from 50–350 mv. in an irregular manner. This is due to the presence of large islands of condensed film in equilibrium with a vapour film of the alcohol. This vapour film has a very small two-dimensional vapour pressure, and a very small surface potential, as found for the vapour of myristic acid.

On compression of the film below about 22 \AA^2 per mol., a slow uniform rise of potential is obtained, but at a definite point B the potential rises suddenly and steeply till the film collapses. This change point agrees within experimental error with the point B on the force/area curve, at $20.0 \pm 0.1 \text{\AA}^2$ per mol.

As in the case of the lower alcohols, the curves on N/10 NaOH are some 50 mv. lower in potential than on N/100 HCl, but the form of the curve is similar. The film appears to be slightly contracted on a strongly alkaline solution; on careful compression an area per molecule as low as 18.6 \AA^2 may be obtained before the film collapses, the whole curve being slightly displaced to the right as compared with that on the acid solution.

Fig. 8 gives the results for eicosyl alcohol. At the lowest temperature the high potential form is not obtained, owing to film collapse, but it appears as a stable form on raising the temperature in exactly the same manner as the high pressure form on the force/area curves (of fig. 6). At temperatures of about

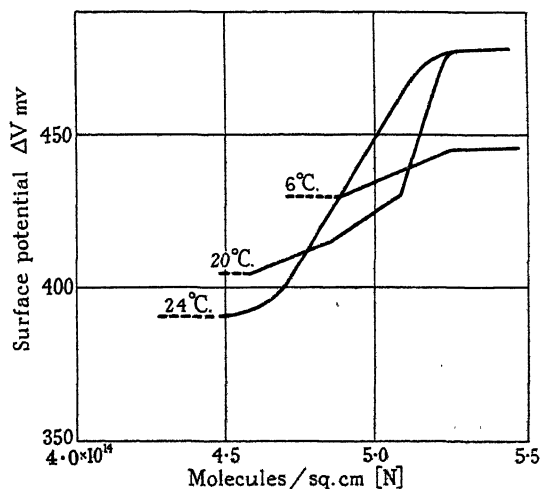


FIG. 8.—Eicosyl alcohol $[C_{20}]$ N/100 HCl.

20° C. the initial portion AB, previously linear, shows a small kink of increasing potential over the same range of area as the arrest level obtained on the force/area curves at the same temperature. The film then passes over into the high potential form.

At higher temperatures the high potential form is obtained at still larger areas, and the same general expansion with temperature as observed on the force/area curves.

DISCUSSION.

In a general way it may be said that the values of the lateral surface pressures indicate the behaviour of the hydrocarbon chain, while the surface potentials indicate that of the polar head group, this being the seat of the predominant electrical dipole in the molecule.

It appears that four factors can operate in determining the structure of the film :—

- (i) The length of the hydrocarbon chain.
- (ii) Temperature.
- (iii) The nature of the underlying solution.
- (iv) The nature of the polar head group.

It is well known from the work of Adam (*loc. cit.*) that increasing the chain length has the same effect as decreasing the temperature; in either case the cohesion of the chains to each other is increased. Thus a film of myristic acid, with 14 carbon atoms at 20° C. behaves similarly, *ceteris paribus*, to a film of palmitic acid with 16 carbon atoms at 40° C.

It is clear, however, that in the general case the nature of the polar head and of the underlying solution must be considered.

The gradual expansion of a film of palmitic acid on increasing the acidity of the underlying solution has already been noticed. The present work shows that this is a particular case of a more general behaviour.

Thus one finds that an alcohol behaves on slightly acid solutions in a similar way to the acid with two more carbon atoms. Myristic acid and dodecyl alcohol both give expanded films, palmitic acid and tetradecyl alcohol both give condensed films, and in all these cases the limiting area increases with increasing acidity.

On concentrated HCl solutions the alcohols are affected much more markedly than the acids. Dodecyl alcohol is raised above its critical point, and changes from an expanded film to a vapour expanded film, and tetradecyl alcohol from a liquid condensed film to an expanded film, while in the case of myristic acid the film remains in the same state as an expanded film, but the limiting area is considerably increased, from 45 A² per mol. (N/100 HCl) to 55 A² per mol. (4 N HCl).

With regard to the condensed films it was at one time thought that the two branches of the typical force/area curve (*e.g.*, AB, BC, fig. 5) could be identified with a "liquid" condensed and a "solid" condensed film respectively.

This view is contradicted by the particular case in point, hexadecyl alcohol, where the condensed film is definitely liquid throughout.

As noticed by Adam (*loc. cit.*, p. 60), palmitic acid and hexadecyl alcohol, under precisely the same conditions (on water at 15° C.), give a solid condensed and a liquid condensed film respectively along BC. This shows that the head group exerts a considerable influence on the behaviour of the hydrocarbon chain for the same chain length, the same temperature and the same underlying solution.

It is thus incorrect to describe the two types of condensed film (AB, BC) as solid and liquid as a comprehensive nomenclature. They will here be designated for convenience by " α " and " β " condensed film respectively.

A second view that has been put forward to explain these two states is that the " α " form represents compression of the head-groups and the " β " form

the compression of the close-packed hydrocarbon chains. It is assumed that the polar head group occupies a larger area of surface, and is more easily compressible than the non-polar chain (Adam, *loc. cit.*). From a study of the force/area curves alone this conclusion is justifiable, but the results obtained on surface potentials indicate that a reconsideration of the case is necessary.

In the " α " form the dipole moment $\bar{\mu}$, derived from the slope of the $\Delta V/n$ curve, is small, approximately $1.1 \cdot 10^{-19}$ e.s.u. per mol. for the alcohols on acid or alkaline solutions, and approximately $1.5 \cdot 10^{-19}$ e.s.u. per mol. for the acids (Part III). That is, the moment of the head group itself is not changing appreciably on compression of the film, as would be expected if the head groups themselves were being compressed. The factor which is altering, however, is the contribution to the total dipole moment from the underlying solution, the term $b/4\pi n$ in equation (3), where b is now positive. This contribution is decreasing on compression of the film, a simple interpretation of which is that the solvent molecules oriented between the polar head groups are being expelled into the underlying solution.

On further careful compression of the film into the " β " form there is a very marked change in the $\Delta V/n$ curves, both in the case of the acids (Part III) and of the alcohols. The potential rises sharply over a very small range of area, the portion BC in fig. 7, prior to film collapse.

This change must be associated with a re-orientation of the CH_2OH or COOH groups when they are brought into close proximity. This re-orientation is assisted by a rise of temperature, occurring at larger areas as the temperature is raised.

As noted by Adam and Dyer (*loc. cit.*), as the chain length is increased the condensed films show a slightly closer packing. This observation is confirmed in this work on the alcohols. For hexadecyl alcohol at 15°C . the transition point B, between the " α " and " β " condensed states occurs at 19.8 \AA^2 ; for eicosyl alcohol this has decreased to 19.5 \AA^2 .

At the same time it must be noticed that the total potential in the condensed films is ca. 40 mv. higher in the case of eicosyl alcohol than for hexadecyl alcohol. No simple explanation of this phenomenon seems available.

The authors are grateful to Dr. N. K. Adam for the gift of pure specimens of myristic acid, hexadecyl alcohol and eicosyl alcohol. They also wish to express their thanks to the Department of Scientific and Industrial Research (J. H. S. and A. H. H.), the Governing Body of Trinity Hall, Cambridge, and

The Merchant Taylors' Company (A. H. H.) for grants which have enabled the investigation to be carried out.

Their very great thanks are due to Professor E. K. Rideal, F.R.S., the instigator of this work, for his continued advice and valuable criticism.

SUMMARY

(1) The effect of the underlying solution on the surface potentials of unimolecular films of long chain aliphatic compounds has been examined over a large range of p_H , and for the specific effects of various ions in the underlying solution.

(2) A distinction is drawn between ionisation of the film-forming material and adsorption of ions on the film.

(3) Two-dimensional transitions of state are discussed in detail.

The Behaviour of Electrolytes in Mixed Solvents. Part IV.—The Free Energy of Zinc Chloride in Water-Alcohol Solutions.

By R. T. HAMILTON, Ph.D., and J. A. V. BUTLER, D.Sc., University of Edinburgh.

(Communicated by J. Kendall, F.R.S.—Received June 25, 1932.)

In Part I of this series* were described measurements of the free energy and heat content of hydrogen chloride in a series of water-alcohol mixtures. It appeared to be very desirable to find if the behaviour shown by hydrogen chloride is typical of that of strong electrolytes in general. For this purpose an attempt was made in the first place to determine the free energies of lithium chloride in these solvents, using amalgam electrodes. Preliminary measurements showed that even under rigidly air-free conditions reproducible values of the electromotive forces of cells containing lithium amalgams could not be obtained in alcohol solutions. Similar experiences with sodium amalgams have been reported from the Balliol College Laboratory,† while Scatchard‡ has also experienced similar difficulties with calcium amalgams. It appeared that

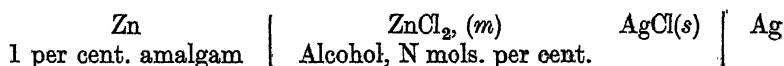
* Butler and Robertson, 'Proc. Roy. Soc.,' A, vol. 125, p. 694 (1929).

† Wolfenden, Wright, Rose Kune, and Buckley, 'Trans. Faraday Soc.,' vol. 23, p. 491 (1927).

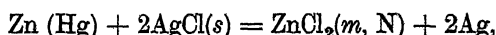
‡ Private communication.

zinc chloride might be suitable, for when proper precautions are taken for the exclusion of air easily reproducible potential differences have been obtained with zinc amalgam in aqueous solutions.* The conductivities† show that zinc chloride is a strong electrolyte in aqueous solution, and although several investigators have found that their product hydrolysed to give turbid solutions in water, we had no difficulty in obtaining a salt which gave perfectly clear solutions at all dilutions. The conductivity measurements of Getman and Gibbons,‡ which show that in alcoholic solution zinc chloride is an extremely weak electrolyte, escaped our notice at the beginning of this work. But although on this account the behaviour of zinc chloride does not extend our knowledge of the properties of completely ionised substances in mixed solvents, it has proved to be an interesting case showing in the range of solvents investigated the transition from a strong to an exceedingly weak electrolyte.

The free energies were determined by measuring the electromotive force of cells of the type



The free energy change in the cell reaction, viz.,



is given by

$$\Delta F = -2EF = -46148E \text{ calories,}$$

where E is the electromotive force and F the electrochemical equivalent. The electromotive forces were determined in solvents containing 0, 25, 50, 90 and 100 mols. per cent. alcohol, each for a range of salt concentrations from about $m = 0.01$ to $m = 1.5$ at $15 \pm 0.02^\circ$. It was originally intended to make a similar series of measurements also at 25° , in order to evaluate the heat content changes, but the reproducibility of the results in solutions containing large proportions of alcohol did not warrant this extension.

Experimental.

Materials.—Ethyl alcohol was prepared by the method described in Part I. Anhydrous zinc chloride was prepared by dissolving A.R. zinc, which was shown to be free from cadmium and iron, in dry ether which was kept saturated

* Richards and Forbes, 'Carnegie Instit. Pub.,' No. 56 (1906).

† Kohlrausch and Holborn, "Leitvermögen der Electrolyte," pp. 147, 168 (1916); Heydweiller, 'Ann. Physik,' vol. 30, p. 873 (1909).

‡ 'Amer. Chem. J.,' vol. 48, p. 125 (1912).

§ m is the number of gram molecules of zinc chloride to 1,000 gm. solvent.

with hydrogen chloride by the passage of the dried gas. When the zinc had completely dissolved the ether was evaporated under reduced pressure, and the zinc chloride finally gently heated *in vacuo*. The product dissolved in water, alcohol and ether to give perfectly clear solutions. Its analysis gave Cl, 51.8 per cent. (theoretical 52.0 per cent.). Zinc amalgams were prepared from mercury which had been well washed with dilute nitric acid and twice distilled at reduced pressure in a current of air, and from electrolytic zinc obtained at platinum cathodes by electrolysis of an ammonical solution of A.R. zinc sulphate. The amalgams were made and transferred to a pipette in an atmosphere of hydrogen by the method described by Richards and Forbes (*loc. cit.*, p. 18). The pipette was sealed at the top and had at the bottom a tap and a long fine delivery tube through which the amalgam could be introduced into the cells. The tip of this tube was normally kept closed with wax, and a part of the amalgam which might have come into contact with air was rejected each time the amalgam was used. The amalgam remained perfectly bright and showed no signs of oxidation at the end of the investigation. Silver-silver chloride electrodes were prepared by a similar method to that employed in Part I, but the silver was plated from an ammoniacal solution instead of from silver potassium cyanide. It is easier to obtain reproducible electrodes in this way.

Electrode Vessels and Technique.—In order to obtain reproducible measurements with zinc amalgam electrodes, it is essential to exclude oxygen from the cell. Since hydrogen causes disturbances with silver chloride electrodes, nitrogen was employed. It was obtained from a cylinder, passed over heated copper filings and dried.

The cell is shown in fig. 1A. The side tubes A, A contain the silver chloride electrodes and the cups, C, C the amalgam. Platinum wires, sealed through the bottom of these cups enabled contact to be made with the leads, which passed down the tubes L, L. Nitrogen entered the cell by the tube E, and could leave when the other openings were closed to the tube F, to which was attached a trap to prevent the entry of air. The tube K, closed by a tight rubber stopper, was for the introduction of solution and amalgam.

The solutions were freed from oxygen in pipettes similar to Fig. 1, *b*, by alternately pumping out and filling with pure nitrogen. In order to fill the cell, the tubes A, A and K were stoppered and a stream of nitrogen passed, care being taken to sweep out all the "dead-ends" with the gas. The silver chloride electrodes were then introduced by removing (one at a time) the stoppers of the tubes A, A and slipping in the electrode while a rapid stream of

nitrogen passed down the open tube to prevent the entry of air. The cups C, C were then filled with the amalgam by similarly introducing the delivery tube of the amalgam pipette through the opening K, and the solution was introduced in the same way.

The electromotive forces of a series of salt concentrations in a given solvent were determined with each set of electrodes. Usually we started with a dilute solution, and when concordant and constant values of the electromotive force

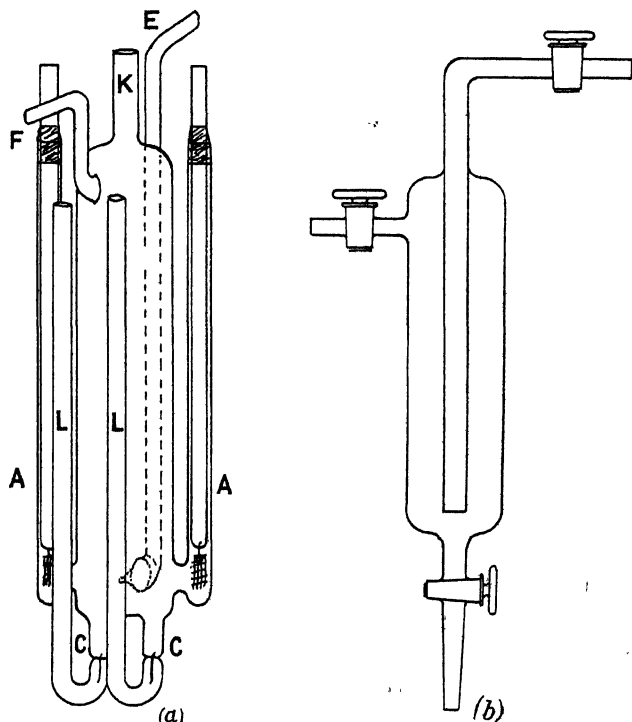


FIG. 1.

had been obtained, we added a quantity of a concentrated solution in the same solvent. Occasionally the most concentrated solution was first placed in the cell and successively diluted. Two such runs were generally required to cover the whole range of salt concentrations. After each determination a sample of the solution was withdrawn and weighed and its salt content determined by precipitating the chloride with a small excess of silver nitrate, filtering and back titrating with standard thiocyanate solution. The maximum error of the analysis was ± 0.1 per cent.

In freeing the solutions from oxygen it is possible that small changes of the solvent composition occurred owing to the unequal evaporation of water and

alcohol. Care was taken to permit no more evaporation than was unavoidable, and it is thought that any such changes as occurred are insignificant.

The electromotive forces were measured with a Tinsley vernier potentiometer, and the usual precautions were taken to avoid thermoelectric effects. The Weston cell used was compared with one which had recently been standardised at the National Physical Laboratory.

Experimental Data.

As an example of the concordance and constancy of the electromotive forces, Table I gives the values obtained consecutively with the two pairs of electrodes in part of a typical run, in which the concentration of the salt was first increased and then decreased.

Table I.

Cell: Zn. 1 per cent. amalgam | $\text{ZnCl}_2(m)$, 25 mols. per cent. alcohol, $\text{AgCl}(s)$ | Ag.

Time (hours).	Electromotive forces.		
	Electrodes 1-3.	Electrodes 2-4.	Mean E.
0	Concentration changed to $m = 0.4925$.		1.0041
2.5	1.00404	1.00426	
3.0	1.00402	1.00422	
3.5	1.00400	1.00418	
4.0	Concentration changed to $m = 0.7447$ (left overnight).		0.9965
22.0	0.99644	0.99658	
22.30	0.99641	0.99650	
24.0	Concentration changed to $m = 0.4105$.		
26.0	1.00661	1.00646	1.0068
26.5	1.00682	1.00676	
27.0	1.00684	1.00680	
27.5	1.00684	1.00680	

Table II gives the average values of the final readings of all the solutions. The values obtained in different runs are usually in good agreement with each other where the concentration ranges overlap and the average error for solutions above $m = 0.05$ is probably within ± 0.0005 volt. Reliable values could not be obtained for salt concentrations less than $m = 0.01$ and there may be errors of ± 0.002 volt in the region of this concentration. Measurements in 90 and 100 mols. per cent. alcohol are difficult on account of the high resistance of the solutions. No values have been recorded which can be directly compared with our results, but as far as a comparison is possible they seem to be in

reasonable agreement with the measurements in aqueous zinc chloride solutions at 25° of Horsch,* and Scatchard and Tefft.†

Table II.

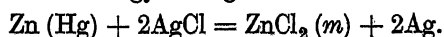
Electromotive forces in volts of the cell

Zn ZnCl₂ (m), AgCl; Ag.
1 per cent. amalgam; Water-alcohol.

Composition of solvent (mol. per cent. alcohol).									
0.		25.		50.		90.		100.	
m.	E.	m.	E.	m.	E.	m.	E.	m.	E.
0.009625	1.1568	0.003490	1.1157	0.002115	1.0839	0.02555	1.0167	0.01263	1.007
0.03040	1.1180	0.006654	1.0940	0.02451	1.0431	0.04244	1.0061	0.02997	1.001
0.1011	1.0799	0.006881	1.1004	0.04656	1.0314	0.08158	0.9998	0.08820	0.9888
0.1031	1.0793	0.01218	1.0813	0.09199	1.0193	0.1000	0.9974	0.2176	0.9783
0.1444	1.0678	0.02611	1.0634	0.1331	1.0177	0.1703	0.9909	0.3178	0.9745
0.3590	1.0387	0.07131	1.0396	0.2479	1.0061	0.4033	0.9766	0.4616	0.9653
0.5285	1.0262	0.2810	1.0139	0.4326	0.9988	0.5989	0.9705	1.047	0.9521
1.202	1.0007	0.4105	1.0068	0.4475	0.9981	0.8044	0.9649		
		0.4925	1.0041	0.7099	0.9894				
		0.7447	0.9965	1.1950	0.9788				
				1.4490	0.9722				

In Table III are given the free energy changes in the cell reaction. The values of E employed in calculating these figures were obtained from graphs of E against m. The free energy change in the transfer of ZnCl₂ from one solution to another can easily be obtained by subtracting the corresponding values in

Table III.—Free Energy Changes in the Cell Reaction (−ΔF).



m.	Composition of solvent (mols. per cent. alcohol).				
	0.	25.	50.	90.	100.
0.01	53300	50160	49050	47530	46520
0.05	50950	48410	47530	46380	45920
0.10	49840	47650	47070	46030	45570
0.50	47440	46330	45960	44930	44490
1.00	46490	45640	45340	44280	43980

* 'J. Amer. Chem. Soc.,' vol. 41, p. 1787 (1919).

† *Ibid.*, vol. 42, p. 2272 (1930).

this table. Fig. 2, in which points obtained in different runs with a given solvent are distinguished, shows the variation of E with $\log_{10} m$ for each solvent, and fig. 3 shows the variation of $-\Delta F$ with the solvent composition for constant values of m .

Discussion.

A uni-bivalent electrolyte may give rise to several types of behaviour according to the type of its dissociation. (1) If it is completely dissociated as, *e.g.*, $\text{ZnCl}_2 = \text{Zn}^{++} + 2\text{Cl}^-$, since two Faradays of electricity are required to form three gram-ions in solution we shall have as an ideal law :

$$E = E_0 - 3RT/2F \log m,$$

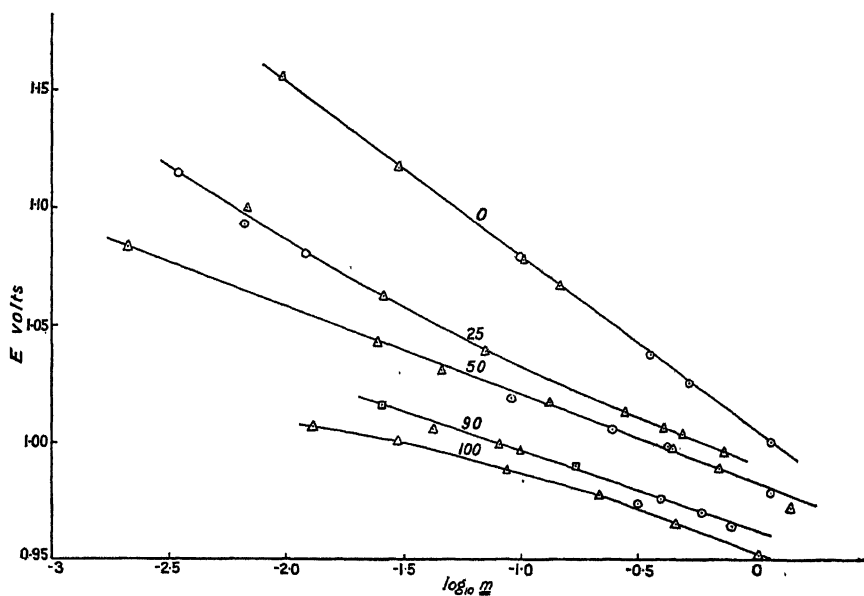
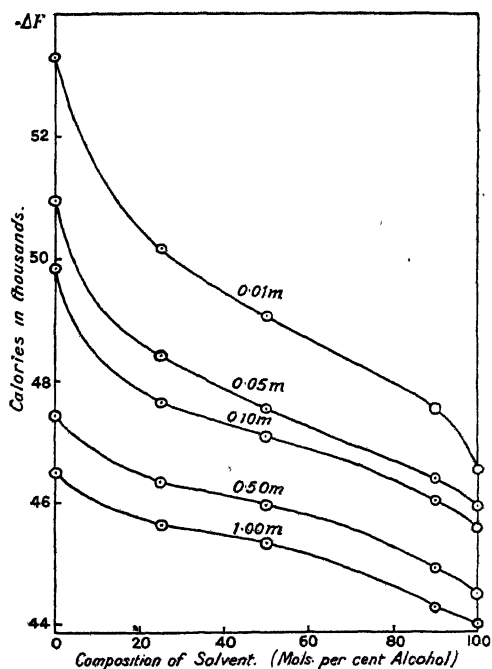
and at 15° the factor $2.303 \times 2RT/2F$ is 0.08575. (2) If the first stage dissociation only, *viz.*, $\text{ZnCl}_2 = \text{ZnCl}^+ + \text{Cl}^-$, is complete the ideal law is $E = E_0 - RT/F \log m$. (3) If it is completely, or practically completely, undissociated, since two Faradays are required to produce 1 molecule in the solution, we shall have $E = E_0 - RT/2F \log m$. The factors $2.303RT/F$ and $2.303RT/2F$ are 0.05717 and 0.02859 at 15° .

The graphs of E against $\log m$, fig. 2, are not far from straight lines over a considerable part of their course, and their slope will give some indication of the kind of dissociation we are dealing with. The average slopes of the curves between $\log_{10} m = -0.5$ and -1.5 are given below :—

Solvent (per cent. alcohol).....	0	25	50	90	100
Average slope of $E - \log_{10} m$...	0.074	0.048	0.038	0.032	0.026

The slope in water is not far from that required for case (1), while that in alcohol indicates complete non-dissociation. This is in accordance with the conductivities (see p. 451). The intermediate solvents have slopes between these extreme values, which do not correspond closely with either case. These are no doubt due to mixed types of ionisation.

It is curious that in water the values of E are almost exactly linear with $\log m$ from $m = 0.01$ to $m = 1$, but the slope differs somewhat from the theoretical slope for an ideal completely dissociated solution. It is possible that the theoretical slope would be obtained at smaller concentrations, but the linearity shown over this extensive range can hardly be accidental. If we write $E = E_0 + k \log mf$, where k is the theoretical slope and f the activity coefficient, and $E = E_0' + k' \log m$, where k' is the empirical slope, we have $\log f = (E_0' - E_0)/k + (k' - k)/k \cdot \log m$, which applies over the range given, but obviously cannot apply down to zero concentration. The value of

FIG. 2.—Electromotive forces plotted against $\log m$.FIG. 3.—Free energy decrease in formation of zinc chloride plotted against N_2 .

$(k' - k)/k$ is about -0.13 . It is difficult to see how such a relation can be reconciled with the form of the activity coefficient obtained from Debye and Huckels' theory.

The differences between the free energies of zinc chloride in water and in alcohol are surprisingly small considering the difference between the state of ionisation in the two solvents. Also, whereas in the case of hydrogen chloride the first additions of alcohol cause a very small change in the free energy, with zinc chloride the first additions of alcohol cause the greatest change. It is important to note that this effect cannot be ascribed to alcohol decreasing the degree of dissociation, for by itself this would cause an increase in the value of $-\Delta F$.

Consider, for simplicity, a binary electrolyte which is completely dissociated in a solvent A, and completely undissociated in a solvent B. Let \bar{F}_s be the partial free energy of the undissociated molecules and \bar{F}_I the sum of the partial free energies of the ions. Then writing, for very dilute solutions,

$$\bar{F}_s = F_s^\circ + RT \log m_s, \quad \bar{F}_I = F_I^\circ + 2RT \log m_I, \quad (1)$$

since it is necessary for internal equilibrium that $\bar{F}_s = \bar{F}_I$, we have

$$\log (m_I)^2/m_s = (F_s^\circ - F_I^\circ)/RT, \quad (2)$$

and $(F_s^\circ - F_I^\circ)/RT = \log K$, where K is the dissociation constant. Now both F_s° and F_I° may be functions of the composition of the solvent. If the salt is practically completely dissociated in the solvent A, then $(F_s^\circ)_A$ must be considerably greater than $(F_I^\circ)_A$ (or $-(F_I^\circ)_A > -(F_s^\circ)_A$), and if the salt is completely undissociated in the solvent B, then $-(F_s^\circ)_B > -(F_I^\circ)_A$.

The points P, Q' and Q, P' on the axis of A and B, fig. 4, represent values of these quantities which are related in this way. Suppose that the values of $-F_I^\circ$, $-F_s^\circ$ in the mixed solvents are represented by the curves PP', QQ'. Knowing these curves it would be easy by (1) and (2) to construct the curve for $\bar{F} = \bar{F}_s$ for a constant value of m . The curve PQ (for $m=1$) will be obtained from values of F_I° , F_s° which vary approximately as PP', QQ'. The case of a uni-bivalent electrolyte such as zinc chloride is somewhat more complicated, but the relations are of a similar form. The general shape of the curves of the free energies of zinc chloride at constant salt concentration can thus be accounted for in terms of varying degrees of dissociation. It will, however, be observed that in order to obtain the initial rapid fall of $-\bar{F}$, it is necessary that $-F_I^\circ$ should also fall rapidly with the first additions of alcohol.

Alternatively, if we know the curve for \bar{F} , and also the dissociation constants in the mixed solvents, it would be possible to deduce the curves PP', QQ'.

We have determined the conductivities of the zinc chloride solutions, which are in accordance with the results of this investigation, but it has proved to be impossible, in the present state of knowledge, to determine from them the degrees

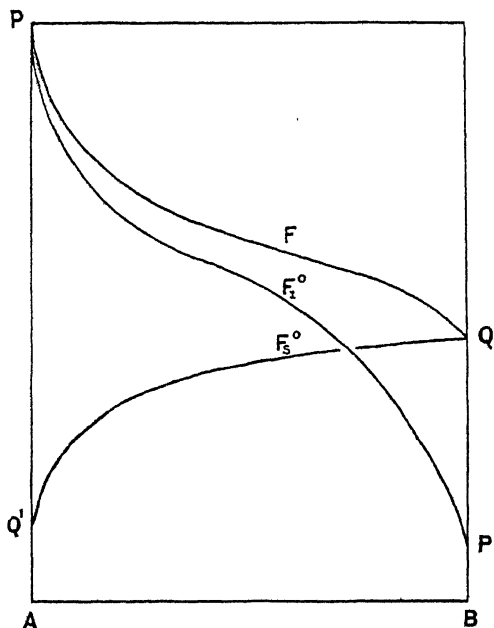


FIG. 4.

of dissociation. Since at present no close correlation between the free energies and the conductivities is possible, we are reserving the measurements of the latter for a more extensive study of the variation of conductivity in mixed solvents.

We desire to express our thanks to the Department of Scientific and Industrial Research for a maintenance grant awarded to one of us (R. T. H.) which made his collaboration possible, and to the Trustees of the Carnegie Universities Trust for a Teaching Fellowship held by J. A. V. B.

Summary.

- (1) The partial free energies of zinc chloride have been determined in a series of water-alcohol solvents at concentrations extending from $m = 0.01$ to $m = 1$.
- (2) The values indicate that zinc chloride is practically completely dissociated in water and practically undissociated in alcohol, which is in accordance with the conductivities of the solutions. The form of the free energy curves is explained in terms of this behaviour.

The Passage of Neutrons through Matter.

By H. S. W. MASSEY, Ph.D., Senior 1851 Exhibitioner, Trinity College,
Cambridge.

(Communicated by R. H. Fowler, F.R.S.—Received August 11, 1932.)

The discovery of the neutron by Chadwick* is of very great importance to the theory of nuclear structure for it is apparently a much simpler unit than the α -particle and so more amenable to mathematical description. It is, therefore, very necessary to obtain as much knowledge as possible of the field of force surrounding the particle. If this field be known a very strict test of any theory of the nature of the particle is provided.

The experimental method used to determine the field of force consists in the observation of the collisions of neutrons with material particles such as electrons and protons. The interpretation of these results requires the development of a satisfactory theory of such collisions. Fortunately, in most cases the smallness of the field of interaction between a neutron and a charged particle leads to the possibility of applying the simple approximate quantum theory of collisions due to Born.† In this paper we will discuss the application of the theory to the elastic collisions of neutrons with material particles. A neutron model consisting of a hydrogen atom in a nearly zero quantum state will be considered in particular and the probability of disintegration of such a model by nuclear impact estimated. It will be shown that the available experimental material indicates that the radius of such an atom must be less than 2.0×10^{-13} cm.

1. Properties of the Neutron and the Atomic Model.

The experiments of Chadwick‡ and Feather§ have shown that the neutron is a particle with mass M nearly equal to that of the hydrogen atom and no appreciable charge. Its velocity of ejection from a beryllium nucleus is roughly 3×10^9 cm. per second and experiments carried out by Dee|| have established that neutrons of such velocity produce less than 1 pair of ions per

* 'Nature,' vol. 129, p. 312 (1932).

† 'Z. Physik,' vol. 38, p. 803 (1926).

‡ 'Proc. Roy. Soc.,' A, vol. 136, p. 692 (1932).

§ 'Proc. Roy. Soc.,' A, vol. 136, p. 709 (1932).

|| 'Proc. Roy. Soc.,' A, vol. 136, p. 727 (1932).

300 cm. path in air at N.T.P. The number of collisions with protons corresponds to a collision radius of the order 10^{-13} cm., while for heavier nuclei the radius effective in collisions is not very different from that of the nucleus concerned. It has also been shown that approximately 25 per cent. of the collisions with nitrogen atoms are inelastic, resulting in disintegration of the nitrogen nucleus. Finally, if one assumes that the neutron is composed of two or more particles its mass shows that the energy of binding of the particle is probably between 1 and 2 million electron volts.

In view of these properties it is of interest to examine the most likely model or the neutron. The fact that its mass is nearly that of the hydrogen atom at once suggests the model of a hydrogen atom in a nearly zero quantum state. Such a possibility was discussed by Rutherford* in 1920 and recently Langer and Rosen† have discussed such a model in terms of the quantum theory. In its present form quantum mechanics excludes this possibility completely, but, since the binding energy involved would be greater than the proper mass of the electron, a relativistic theory of the interaction of two bodies is required. No such theory exists at present. We are, therefore, justified in disregarding the impossibility of the nearly zero quantum state in the present quantum theory and will imagine the neutron as an atom in which the electron moves in a field of very high effective nuclear charge Z . The field of such an atom will be given by

$$V(r) = e^2 \left(\frac{1}{r} + \frac{Z}{a_0} \right) e^{-2Zr/a_0}, \quad (1)$$

where a_0 is the radius of the first Bohr orbit of hydrogen. The radius of the neutron will then be a_0/Z . If we assume that the electron and proton both behave as point charges down to very small distances we may represent collisions of neutrons with these particles as the scattering of particles by the potential field $V(r)$.

2. Theory of Collisions.

The wave equation for the relative motion of the neutron and nucleus is

$$\nabla^2 \psi + \left\{ k^2 - \frac{8\pi^2 M}{h^2} V(r) \right\} \psi = 0, \quad (2)$$

where M is the reduced mass of the system and the wave-number $k = 2\pi Mv/h$, v being the relative velocity of the colliding particles. We require a proper

* 'Proc. Roy. Soc.,' A, vol. 97, p. 374 (1920).

† 'Phys. Rev.,' vol. 37, p. 1579 (1931).

solution of this equation which has the asymptotic form

$$\psi \sim e^{ikx} + f(\theta, \phi) \frac{e^{ikr}}{r}. \quad (3)$$

The collision cross-section is then given by

$$Q = \int_0^{2\pi} \int_0^\pi |f(\theta, \phi)|^2 \sin \theta \, d\theta \, d\phi. \quad (4)$$

To obtain such a solution we expand ψ in the form

$$\psi = \sum_n \psi_n(r) P_n(\cos \theta). \quad (5)$$

The function ψ_n must then satisfy

$$\frac{d^2}{dr^2}(r\psi_n) + \left\{ k^2 - \frac{8\pi^2 M}{\hbar^2} V(r) - \frac{n(n+1)}{r^2} \right\} (r\psi_n) = 0, \quad (6)$$

and will have the asymptotic form

$$\psi_n \sim \frac{A_n}{kr} \sin(kr - \tfrac{1}{2}n\pi + \delta_n). \quad (7)$$

It has then been shown by Faxén and Holtsmark* that the collision cross-section is given by

$$Q = \frac{4\pi}{k^2} \sum_n (2n+1) \sin^2 \delta_n. \quad (8)$$

The problem then reduces to that of calculating the phases δ_n . If the scattering field is small approximate expressions may be used for these. Thus it has been shown by Mott† that δ_n is given approximately by

$$\delta_n = \frac{4\pi^2 M}{\hbar^2} \int_0^\infty r V(r) \{J_{n+\frac{1}{2}}(kr)\}^2 dr, \quad (9)$$

provided the right-hand side is considerably less than unity. In this case the series (8) for Q may be summed by assuming

$$\sin \delta_n = \delta_n$$

and we arrive at Born's formula

$$Q = \frac{8\pi^3 M^2}{\hbar^4} \int_0^\pi \left| \int V(r') e^{2ikr' \cos \theta' \sin \frac{1}{2}\theta} dv' \right|^2 \sin \theta \, d\theta. \quad (10)$$

* 'Z. Physik,' vol. 45, p. 307 (1927).

† 'Proc. Camb. Phil. Soc.,' vol. 25, p. 304 (1928).

The condition of validity of this formula is that the phases be all considerably less than unity. It appears from comparison of exact and approximate values of δ_n for the scattering of electrons by atoms that the right-hand side of (9) does not need to be much less than 0.5 in order that it shall represent a good approximation to δ_n .*

3. Application to Neutron Collisions.

(a) *With Heavy Nuclei.*—If Z' is the nuclear charge, the field of interaction between a neutron and the nucleus, on the basis of the assumptions of section (1) will be

$$V(r) = Z'e^2 \left(\frac{1}{r} + \frac{Z}{a_0} \right) e^{-\frac{2Zr}{a_0}} \quad (11)$$

at distances greater than the nuclear radius. As the experimental evidence shows that the collision diameter is not greater than the nuclear radius we are enabled to find a lower limit to the value of Z .

Using the values of the mass and velocity of the neutron we find

$$k = 4.76 \times 10^{12} \text{ cm.}^{-1},$$

so the collision cross-section

$$Q = 5.5 \times 10^{-25} \sum_n (2n+1) \sin^2 \delta_n.$$

The observed cross-section for lead is

$$\pi \times (7 \times 10^{-13})^2 = 1.54 \times 10^{-24}$$

so

$$\sum_n (2n+1) \sin^2 \delta_n = 2.8.$$

This immediately sets an upper limit to $\sin^2 \delta_n$, which cannot be greater than $2.8/(2n+1)$. In effect, this limits the size of the phase itself for, if δ_1 , say, is greater than $\pi/2$, some higher phase will be nearly as great as $\pi/2$, and will give too large a value for the cross-section. We may then take it that δ_2 cannot be greater than 0.7. Actually it must be very much less than this owing to the additive effect of the other terms and we can assume that all the phases except the zero order one must be sufficiently small to be given by the approximate form (9). We may then write the expression for the cross-section

* J. McDougall, 'Proc. Roy. Soc., A, vol. 136, p. 549 (1932).

in the form

$$Q = \frac{8\pi^3 M^2}{\hbar^4} \int_0^\pi \left| \int_0^\infty V(r') e^{2ikr' \sin \frac{1}{2}\theta \cos \theta'} dv' \right|^2 \sin \theta d\theta \\ - \frac{4\pi}{k^2} \left[\frac{16\pi^4 M^2}{\hbar^4} \left\{ \int_0^\infty r V(r) J_{n+\frac{1}{2}}(kr) dr \right\}^2 - \sin^2 \delta_0 \right], \quad (12)$$

where the second term is a correction to the Born formula (10) for the zero order term.

Substituting the expression (11) for $V(r)$ in the formula (12) we find, on carrying out the integrations

$$Q = \frac{4\pi^5 M^2 e^4 Z'^2}{k^4 \hbar^4} \left[\frac{1}{3} \frac{48x^4 + 72x^2 + 28}{x^2(x^2 + 1)^3} - 4 \left\{ \log \left(1 + \frac{1}{x^2} \right) + \frac{1}{1 + x^2} \right\}^2 \right] \\ + \frac{4\pi}{k^2} \sin^2 \delta_0, \quad (13)$$

where

$$x = Z/ka_0.$$

If we neglect, for the moment, the term in $\sin^2 \delta_0$, the magnitude of which we cannot obtain except by integrating the differential equation (6), we may find a value of x which would lead to a value of Q of the right magnitude for collisions with lead nuclei. This will give a lower limit to Z . In this way we find

$$x > 1.0$$

so

$$Z > 25,000. \quad (14)$$

The approximate expression (9) for δ_0 becomes equal to 3.5 which shows that the correcting term in (12) is important.

(b) *Light Nuclei*.—Let us now consider the consequences of the lower limit obtained above for the size of the neutron.

For light nuclei we see that Born's approximation may be used for all the phases, for the zero order phase becomes

$$\delta_0 = 3.5 \left(\frac{Z'}{80} \right),$$

where Z' is the nuclear charge. The cross-sections of the light nuclei would then be in the ratio of the squares of the nuclear charges. This is in contradiction to the observations which show that the variation with Z' is much slower. It is clear then that the observed collision areas must be determined

by the internal nuclear field, the external field of the neutron being of such short range that the colliding systems penetrate very closely.

(c) *Protons*.—The above considerations become of special importance when applied to protons. If the proton behaves as an elementary charge down to very small distances of approach the collision area will be given by

$$Q \simeq \frac{16\pi^5 M^2 e^4 a_0^4}{h^4 Z^4} \quad (15)$$

where M is the mass of the proton. Using the above lower limit for Z we find that the collision radius must be less than 1.4×10^{-14} cm. All observations seem to indicate a much greater value than this. Consistency with the above conclusions would then require the introduction of a powerful internal proton field coming into play at a distance given roughly by the collision radius associated with neutron-proton collisions.

In view of these conclusions it would seem to be of great interest to obtain accurate measurements of collision cross-sections. Although the above results have been obtained by the use of a special field it is likely that, provided the interaction of a charge with a neutron is proportional to the magnitude of the charge, the results will hold for any field.*

(d) *Electrons*.—For this case $M \simeq m$ the mass of the electron and the collision radius becomes

$$\alpha_{elec} = 8\pi^2 m e^2 a_0^2 h^{-2} Z^{-2},$$

and is $2/1850$ of that for proton-neutron collisions. For $Z = 2.5 \times 10^4$ we find that

$$\alpha_{elec} = 7.4 \times 10^{-17} \text{ cm.}$$

The number of ions produced per centimetre of path by neutrons will thus be very small as may be seen from the following considerations. In passing through air containing 5.3×10^{20} electrons per cubic centimetre the total number of such collisions per centimetre path would be

$$\pi \times 49 \times 10^{-38} \times 5.3 \times 10^{20} = 3 \times 10^{-13} \text{ per cm.}$$

The maximum loss of energy by a neutron per collision with an electron is $2mv^2$, where v is the velocity of the neutron. Allowing 30 volts per ion, the

* As it is unlikely on general grounds that the radius of the proton is nearly as great as 10^{-13} cm. (assuming it to be comparable with e^2/Mc^2) the experimental results would seem to indicate that it is not the nuclear charge which is important in determining the interaction between neutrons and nuclei at large distances. The effective property must be one which varies much more slowly than the nuclear charge in passing from element to element.

total number of ions formed per centimetre path could not be greater than

$$3 \times 10^{-13} \cdot 2mv^2 a_0 e^{-2} = 1 \text{ ion per } 10^{10} \text{ cm.},$$

that is to say, practically no ions at all. If any ions are detected with certainty then it is clear that Z must be considerably smaller than 25,000.

The negative results of Dee's experiments are thus easily explicable. Even if an internal electron field must be taken into account the small mass of the electron will still reduce the cross-section to a low value unless the field be very strong indeed.

4. Probability of Disintegration of a Neutron by Impact with Protons and Heavy Nuclei.

The probability of ejection of an electron from an atom of mass M_1 with velocity between v'' and $v'' + dv''$ by the impact of a charged particle of mass M_2 and charge $Z'e$, moving with initial velocity v relative to the atom, is given by*

$$Q(k'') dk'' = \frac{2^{11} \pi M^2 a_0^2 Z'^2 k'' dk''}{Z^4 m^2 k^2} \int_{k-k'}^{k+k'} \frac{\{K^2 + \frac{1}{2}(1+k''^2)\} K^{-1}}{\{(K+k'')^2 + 1\}^3 \{(K-k'')^2 + 1\}^3} \\ \cdot \exp \left\{ -\frac{2}{k'} \arctan \frac{2k''}{K-k''^2+1} \right\} \left\{ 1 - e^{-\frac{2\pi}{k''}} \right\}^{-1} dK, \quad (16)$$

where

$$k = \frac{2\pi M v a_0}{hZ}, \quad M = \frac{M_1 M_2}{M_1 + M_2}, \quad k'' = \frac{2\pi M v'' a_0}{hZ}, \\ k'^2 = k^2 - \frac{8\pi^2 M a_0^2}{Z^2 h^2} E - \frac{M}{m} k''^2, \quad (17)$$

E is the ionisation energy of the atom and Z is the effective nuclear charge. If, now, we assume, as a rough approximation, that the neutron behaves in this way as an atom with large effective nuclear charge we may use the above formula to estimate the probability of disintegration of a neutron by impact with nuclei.

We take E as corresponding to 1.5×10^6 electron volts and v as 3×10^9 cm. per second. This gives

$$k'^2 = 0.31 - \frac{M}{m} k''^2 \quad \text{for a heavy nucleus} \\ = 0.04 - \frac{M}{2m} k''^2 \quad \text{for a proton,}$$

M being the mass of the neutron.

* Bethe, 'Ann. Physik,' vol. 5, p. 325 (1930).

With these values we may write with sufficient accuracy

$$\int_{k-k'}^{k+k'} \frac{K^2 + \frac{1}{3}(1+k''^2)}{K \{(K+k'')^2 + 1\}^3 \{(K-k'')^2 + 1\}^3} \exp \left\{ -\frac{2}{k''} \arctan \frac{2k''}{(K^2 - k''^2 + 1)} \right\} dK \\ = \frac{k^2 + \frac{1}{3}(1+k''^2)}{k \{(k+k'')^2 + 1\}^3 \{(k-k'')^2 + 1\}^3} \exp \left\{ -\frac{2}{k''} \arctan \frac{2k''}{k^2 - k''^2 + 1} \right\}. \quad (18)$$

The total probability of ionisation will be obtained by integrating over all possible values of k'' , i.e., from $k'' > 0$ to $k'' = \frac{m}{M} \left(k^2 - \frac{8\pi^2 M E a_0^2}{h^2 Z^2} \right)^{\frac{1}{2}}$. Carrying this integration out approximately gives for the probability cross-section corresponding to disintegration of the neutron

$$Q(k) = \pi a_0^2 \frac{2^{10} Z^2 (k^2 + \frac{1}{3})}{Z^4 k^3 (k^2 + 1)^6} e^{-4(1+k^2)} \left(k^2 - \frac{8\pi^2 M E a_0^2}{Z^2 h^2} \right). \quad (19)$$

Taking the value, 2.5×10^4 , for Z gives

$$Q(k) \simeq \pi a_0^2 \times 2 \times 10^{-16} Z^2, \quad (20)$$

corresponding to a collision radius a , for disintegration, $a \simeq 1.70 \times 10^{-16} Z'$ cms. for a heavy nucleus, and a radius of the same order for a proton.

These collision areas are very small and indicate that it is very unlikely that any disintegration of a neutron would be observed in its passage through matter.*

Concluding Remarks.

It has been shown above that the model assumed for the neutron is capable of explaining the experimental results but it is of some interest to examine what effects would be expected on other hypotheses.

Firstly we note that, as long as Born's formula is valid and the wave-length associated with the collision is long compared with the radius of the neutron, collisions with a proton are $(M/m)^2$ times as frequent as with an electron. This follows from the fact that when the wave-length is long, only the zero order cross-section is important and Born's formula gives for this the form

$$Q_0 = \frac{8\pi^3 M^2}{h^4} \left| \int V(r) dv \right|^2 \quad (21)$$

* Throughout this calculation it has been assumed that the out-going electron is moving in the field of a charge $Z = 2.5 \times 10^4$ but actually the density of the outgoing electron near the neutron radius will be less than that which would correspond to such a charge. This will result in less overlap of initial and final electron wave functions and a smaller probability of disintegration than that calculated above.

in the low velocity limit, M being the reduced mass of the colliding systems. As the much greater collision area for proton than for electron collisions is a feature of the experimental results the field of the neutron must vary so gradually with closeness of approach that Born's approximation is valid, or, if the internal fields of the electron and proton are involved, these cannot be so intense that the formula (21) fails to represent the facts approximately.

A possibility which finds favour in some quarters is that the neutron behaves like a dipole. If we assume that the field varies as $\alpha e^2/r^2$ we have from (9) the condition of validity of Born's formula, that

$$\frac{2\pi^2 M \alpha e^2}{h^2 (n + \frac{1}{2})} \ll 1. \quad (22)$$

If, then, the dipole moment is much less than 2.8×10^{-21} c.g.s. units we may apply the usual approximation (10). Using this we find for the distribution per unit angle of protons struck, the form

$$I(\phi) \sin 2\phi \, d\phi = \frac{2\pi^5 \alpha^2 e^4 h^2}{v^2} \sec^2 \phi \sin 2\phi \, d\phi. \quad (23)$$

It should be readily possible to distinguish such a distribution from that which would be obtained if the interaction falls off more rapidly than r^{-2} for large r . In the latter case the number scattered per unit angle tends to zero as the angle tends to zero or to $\pi/2$, whereas with a dipole the number tends to infinity as the angle tends to $\pi/2$.

In particular with the field (1) the distribution per unit angle of struck protons is given by

$$I(\phi) \sin 2\phi \, d\phi = \frac{2\pi^5 M^2 e^4 a_0^4}{h^4} \frac{(2Z^2 + k^2 a_0^2 \cos^2 \phi)^2}{(Z^2 + k^2 a_0^2 \cos^2 \phi)^4} \sin 2\phi \, d\phi. \quad (24)$$

Comparison with the results of Wilson Cloud chamber observations should help to test the validity of this formula.

I wish to express my indebtedness to Dr. J. Chadwick for much discussion of the observational material.

Summary.

The quantum theory of collisions due to Born and Faxén and Holtsmark is applied to the consideration of the collisions of neutrons with atomic nuclei and with electrons. A neutron is represented as an atom in a nearly zero quantum state and it is shown that the experimental results indicate that the

radius of such an atom must be less than 2.0×10^{-13} cm. Evidence is deduced that, either the proton does not behave as a unit charge for distances of approach less than 10^{-13} cm., or the energy of interaction between neutrons and nuclei at large distances increases much less rapidly than the nuclear charge in passing from element to element.

Further, it is shown that the probability of disintegration of a neutron by nuclear impact is very small if the assumed model is valid.

The Large Angle Scattering of Electrons in Gases.—II.

By C. B. O. MOHR, B.A., M.Sc., Trinity College, Cambridge, 1851 Exhibitioner, University of Melbourne, and F. H. NICOLL, B.Sc., M.Sc., Trinity College, Cambridge, 1851 Exhibitioner, University of Saskatchewan.

(Communicated by Lord Rutherford, O.M., F.R.S.—Received August 20, 1932.)

Introduction.

The results of the investigation of the angular distributions of electrons scattered elastically and inelastically in helium, argon, and mercury vapour have been described in a previous paper.* It was found that the inelastic angular distributions in mercury vapour and argon exhibited diffraction maxima and minima similar to those which appeared in the elastic angular distributions.†

Recently Tate and Palmer‡ have investigated the elastic scattering in mercury vapour for angles of scattering between 5° and 125° , and the inelastic scattering (corresponding to a discrete energy loss) between 5° and 55° . Hughes and McMillen§ have measured the angular distributions of the elastic and inelastic scattering in hydrogen between 40° and 165° , and the elastic scattering in helium between 15° and 150° . These papers are also concerned with the large angle scattering of those electrons which make ionising collisions.

The present paper describes the continuation of our work and gives angular distributions for hydrogen, carbon dioxide, methane, nitrogen, neon, phosphine, and hydrogen sulphide. The elastic and inelastic scattering curves are again found to be similar, and the results taken as a whole greatly clarify the interpretation of the phenomena.

* 'Proc. Roy. Soc.,' A, vol. 138, p. 229 (1932), referred to throughout as Paper I.

† Bullard and Massey, 'Proc. Roy. Soc.,' A, vol. 130, p. 579 (1931); vol. 133, p. 637 (1931). Arnot, 'Proc. Roy. Soc.,' A, vol. 130, p. 655 (1931); vol. 133, p. 615 (1931).

‡ 'Phys. Rev.,' vol. 40, p. 731 (1932).

§ 'Phys. Rev.,' vol. 41, pp. 39 and 154 (1932).

Method.

The apparatus used in these experiments was the same as that described in Paper I and it will be sufficient to mention briefly the essential parts. The source of electrons was a gun consisting of a thoriated tungsten filament enclosed in a case containing two slits for defining the electron beam. The case was maintained at a suitable potential for producing electrons of the desired velocity. The electron gun could be rotated about the centre of a field-free scattering space by means of a ground glass joint, in such a way that the beam of electrons crossed the centre of this space. Electrons scattered in a small region near the centre passed through two defining slits into an electrostatic analyser, consisting of two curved plates, and then into the Faraday cylinder. The potential on these plates was adjusted to such a value that only electrons of the desired velocity were collected by the Faraday cylinder. The whole system was enclosed in a pyrex vessel and with the exception of the ground glass joint, could be baked at a temperature of 450°C . The pressure of the gas in the apparatus was usually maintained at a steady value of less than 6×10^{-4} mm. Hg by means of a constant flow method, the gas being admitted to the apparatus through a fine leak and then pumped away. Liquid air traps on the inlet and outlet tubes excluded mercury vapour from the apparatus. All the gases used with the exception of phosphine and hydrogen sulphide, were obtained from commercial cylinders, and apart from being dried were not further purified. The phosphine and hydrogen sulphide were prepared chemically as described later.

Results.

The curves obtained for the elastic scattering of electrons in hydrogen, methane, nitrogen, and neon agree closely in form with those obtained by Arnot (*loc. cit.*) for angles of scattering less than 120° . In the present work the readings were extended to an angle of about 160° . The elastic scattering was also investigated in carbon dioxide, phosphine, and hydrogen sulphide. The angular distributions of electrons which have lost energy in exciting the atoms or molecules to the first excited level were measured in hydrogen, carbon dioxide, nitrogen and neon. The intensity of the inelastic scattering at large angles in the first three gases was usually between one-tenth and one-twentieth of the intensity of the elastic scattering, while in neon it was about one-fiftieth. The ratio was greater at the higher voltages studied and became less at the lower voltages; below about 40 volts the small intensity of the inelastic scattering at large angles prevented measurements being taken. For the

molecules, methane, phosphine, and hydrogen sulphide, no well-marked discrete energy losses could be found, although a large amount of ionisation was observed.

(a) *Hydrogen*.—Angular distributions of 63 and 125 volt electrons scattered elastically and inelastically are shown in fig. 1. The main discrete energy

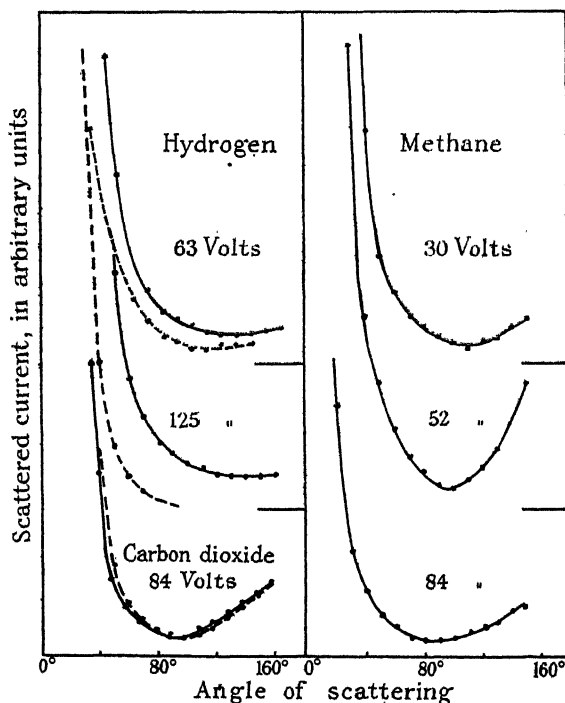


FIG. 1.—Scattering in hydrogen, carbon dioxide, and methane. — Elastic. ... Inelastic.

loss* is 12.6 volts, and corresponds almost certainly to excitation of the molecule to the C state ($2^1\Pi$ level). For 63-volt electrons, the elastic and inelastic angular distributions are very similar and both show a slight rise at large angles. The results agree with those of Hughes and McMillen for 50-volt electrons, although they find a more pronounced rise at large angles. At 125 volts the elastic curve becomes flat at angles greater than 120° , while the inelastic curve falls off very rapidly, the intensity becoming too small to be measured beyond 80° . Again, for 100-volt electrons Hughes and McMillen obtain a slight rise in the elastic angular distribution at large angles, a feature which is not present in our 125-volt curve. As was pointed out in Paper I, the apparatus was primarily designed for investigating inelastic scattering at

* Whiddington, 'Phil. Mag.', vol. 6, p. 889 (1928); Harnwell, 'Phys. Rev.', vol. 35, p. 285 (1930); Hughes and McMillen, 'Phys. Rev.', vol. 36, p. 1034 (1930).

large angles where the intensity is very small, and as a result the measurements cannot be regarded as very accurate at angles less than about 40° .*

(b) *Carbon Dioxide and Methane*.—The angular distributions of the elastic scattering of 84-volt electrons in carbon dioxide and 30, 52 and 84-volt electrons in methane are illustrated in fig. 1. These curves all show a very pronounced minimum at about 90° . An energy loss of 12.9 volts occurs in carbon dioxide,† the interpretation of which is doubtful. The angular distribution of electrons having suffered this loss is also illustrated in fig. 1 for incident electrons of 84 volts energy, the curve being arbitrarily fitted to the corresponding elastic curve at the minimum.

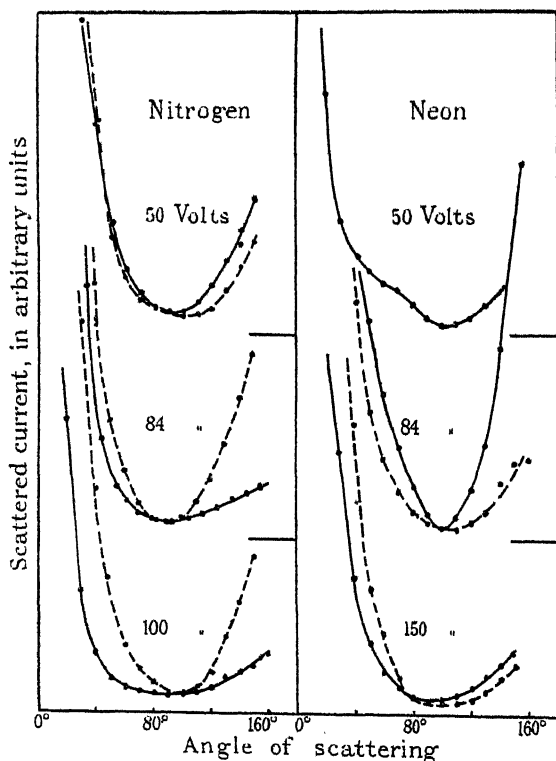


FIG. 2.—Scattering in nitrogen and neon. — Elastic. ... Inelastic.

* *Note added in proof, September 26, 1932.*—In Paper I, this reason was advanced to explain the unexpected lack of agreement at small angles of the experimental curves or the elastic scattering of 100-200 volt electrons in helium, with the theoretical curves obtained by the use of the Born formula. The experimental curves, however, agree quite well with those recently obtained by Hughes, McMillen and Webb (*loc. cit.*) for helium. The discrepancy with theory is probably due to the neglect of the effect of polarisation.

† Rudberg, 'Proc. Roy. Soc.,' A, vol. 130, p. 182 (1931).

(c) *Nitrogen and Neon*.—The angular distributions of electrons scattered elastically and inelastically in nitrogen and neon are shown in fig. 2, the curves being fitted together arbitrarily at the minimum. The nitrogen curves are for 50, 84, and 100-volt electrons and the neon curves for 50, 84, and 150-volt electrons. The inelastic loss investigated in nitrogen* was the 12·8 volt loss corresponding to excitation to the first singlet level of the molecule, while the loss investigated in neon† was the 16·6 volt loss due to excitation to the 2^1P level. The elastic and inelastic curves are again seen to be similar as in the case of carbon dioxide, while all the curves exhibit the pronounced minimum which also appears in the curves for carbon dioxide and methane.

(d) *Phosphine, Hydrogen Sulphide, and Argon*.—The angular distributions of the elastic scattering in these three gases are compared, in fig. 3, for electrons

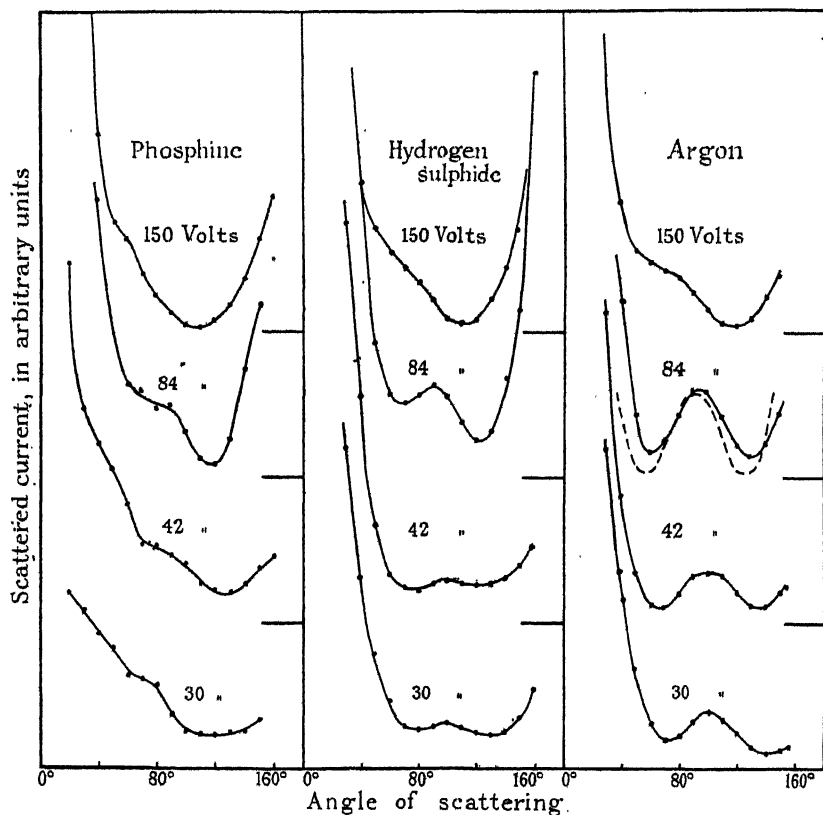


FIG. 3.—Elastic scattering in phosphine, hydrogen sulphide and argon. ...Form of the curve $|P_2(\cos \theta)|^2$.

* Rudberg, 'Proc. Roy. Soc.,' A, vol. 129, p. 628 (1930).

† Van Atta, 'Phys. Rev.,' vol. 38, p. 876 (1931).

of 150, 84, and 30 volts energy respectively. Their interest lies both in the general similarity of the curves and their gradual change from phosphine to hydrogen sulphide, and to argon. The positions of the maxima and minima are almost the same in the three gases, but the prominence of the peak at about 100° increases as one goes from phosphine to argon. Since the scattering at the higher voltages by phosphine and hydrogen sulphide is probably mainly due to the phosphorus and sulphur atoms, the curves therefore establish the similarity of the scattering by elements near each other in the same row of the periodic table.

The phosphine used in these experiments was prepared by dropping a solution of potassium hydroxide in methylated spirits on to phosphorus *in vacuo*. The gas obtained in this way was purified by fractional distillation from a liquid air trap. The hydrogen sulphide was obtained by dropping sulphuric acid on iron sulphide *in vacuo*. An attempt was made to investigate the scattering in hydrogen chloride and chlorine, since chlorine lies between sulphur and argon in the periodic table. The results, however, were not considered reliable owing to the strongly corrosive action of these gases on the tap grease and various parts of the apparatus.

Discussion.

A comparison of the elastic and inelastic angular distributions at large angles in nitrogen, carbon dioxide and neon shows the same close similarity which was shown in Paper I to exist in the case of argon and mercury vapour. This similarity was discussed in Paper I, and it was suggested that the results provided a clue to certain theoretical difficulties. It was predicted that, although the scattering of electrons by atoms is a process involving many complicating effects due to polarisation, electron exchange, and the distortion of the incident and outgoing electron waves by the field of the atom, the observed results might be explained on simple considerations involving only the "size"* of the atom. The experimental results given in the present paper indicate how all the observed facts can in fact be explained qualitatively on the basis of a simple picture.

Consider first the elastic scattering. For electrons of any given velocity the angular distribution may be expressed in the form

$$|\Sigma a_n P_n(\cos \theta)|^2, \quad (1)$$

* If the atom is pictured as a spherical hole, then the term "size" is considered to include not only the width of the hole but also its depth.

where θ is the angle of scattering, the a_n are complex numbers depending on the velocity of the incident electrons and the atomic field, and where as usual

$$\left. \begin{aligned} P_0(\cos \theta) &= 1 \\ P_1(\cos \theta) &= \cos \theta \\ P_2(\cos \theta) &= \frac{1}{2}(3 \cos^2 \theta - 1), \text{ etc.} \end{aligned} \right\}. \quad (2)$$

At sufficiently high velocities, the approximation due to Born* can be used, in which the incident electron wave is regarded as plane, and this leads to values of the coefficients a_n in expression (1) which give merely a monotonic fall in the scattered intensity with increasing angle of scattering. At lower velocities it is necessary to include, among other things, the effect of the distortion of the incident and outgoing electron waves by the field of the atom. This will have the effect of altering the coefficient of one or more of the spherical harmonics in expression (1). Now the scattered intensity given by the Born formula falls away to quite small values at large angles of scattering (except at quite low velocities), and if it is the coefficient of the n^{th} harmonic which is mainly affected by the distortion of the electron waves, then the angular distribution of the elastic scattering will be given approximately by $|P_n(\cos \theta)|^2$ at large angles. If we consider the scattering of electrons of a given velocity by different atoms, it is easily shown that for the larger atoms it will be the harmonics of higher order which will be important, while for the lighter atoms only the harmonics of lower order will be important.

We will now consider the experimental results for the elastic scattering. The curves in this paper for hydrogen† and in Paper I for helium are on the whole nearly flat at angles greater than about 90° . Referring to the values of the spherical harmonics given in (2), we see that for these light elements, it is chiefly the harmonic P_0 which is effective. When we come to the curves for nitrogen and neon, it is seen that the curves are roughly of the form $\cos^2 \theta$. Referring again to (2), this indicates that, for the elements at the end of the second row of the periodic table in which the addition of the L shell of electrons is being completed, it is chiefly the first order wave which is distorted, involving the harmonic P_1 .

We now come to the curves given in fig. 3 for phosphine, hydrogen sulphide, and argon, and in comparing them we expect that at the velocities in question

* 'Z. Physik,' vol. 38, p. 803 (1926).

† The scattering by a homonuclear diatomic molecule has been shown (Massey and Mohr, 'Proc. Roy. Soc.,' A, vol. 135, p. 258 (1932)) to be practically the same as that caused by the two separate atoms.

the curves for phosphine and hydrogen sulphide are well representative of the scattering by the phosphorus and sulphur atoms respectively. Now the curves for argon are approximately of the form* $[P_2 (\cos \theta)]^2$, so that in the case of argon it is the second order wave which is mainly distorted. The curves for phosphine and hydrogen sulphide are similar in form, but are intermediate in character between the curves for neon and argon, so that in the case of phosphorus and sulphur, the first order wave is being distorted almost as much as the second order wave. Actually, of course, some of the other harmonics are also affected, and to obtain quantitative agreement with experiment, it is necessary to calculate a number of the coefficients a_n of the slowly convergent series in (1) using the method of Faxen and Holtsmark.†

Now in the above discussion it has been assumed that the velocity of the incident electrons is sufficiently high for the scattering to be mainly due to the inner shells of the atom—as is the case with the results given in this paper—and not merely to the outer shell of electrons. If this is the case, the precise nature of the outer shell is of secondary importance as far as the scattering at large angles is concerned. For example, the curves for hydrogen sulphide are similar to those for argon, so that the effect of the two hydrogen atoms on the scattering is not very great at the voltages considered. Similarly we may expect that the scattering by an excited atom in which an outer electron is removed to a greater distance from the nucleus will be similar to the scattering by the normal atom at not too low velocities.‡ Hence in the distortion of the outgoing electron wave which has lost energy in exciting the atom, the same harmonic is disturbed as in the distortion of the incident electron wave, provided that the velocity of the incident electrons is large compared to the energy lost in exciting the atom. Hence we expect the inelastic angular distributions to be closely similar to the elastic at large angles and high enough velocities. This is in fact what is observed.

Let us now turn to the effect of the exchange of electrons between the atom and the colliding beam. Since the outgoing exchanged electrons have the same energy as, and move in the same field as, the outgoing non-exchanged electrons, application of the above considerations shows that the exchanged electrons will have just the same angular distribution as the non-exchanged

* The form of this function is illustrated by a dotted line in fig. 3 for purposes of comparison.

† 'Z. Physik,' vol. 45, p. 307 (1927).

‡ The foregoing arguments will not apply to the consideration of the total cross sections which is mainly due to the scattering at small angles.

electrons at large angles of scattering. This shows why it has been possible to explain the angular distributions by entirely neglecting the effect of exchange. Also the simple considerations of this section lead one to expect that a knowledge of the precise form of the atomic field may not be necessary to obtain good agreement with experiment. Thus Morse and Allis* were able to obtain quite good agreement with experiment with the use of a comparatively crude model of the atomic field. Again, it has been shown possible† to explain the observed angular distributions in mercury vapour by using the Fermi field together with an approximate method of calculating the coefficients a_n in expression (1). We thus see, as was suggested in Paper I, that it is merely the "size" of the atom which is of fundamental importance in the consideration of the scattering of electrons by atoms.

At lower velocities the simple considerations of this section no longer apply. For the heavier atoms, only the outer shells will be effective in the scattering, and further the wave-lengths of the incident electrons and the inelastically scattered electrons will differ by a greater amount, so that the inelastic angular distributions will become less similar to the elastic at lower voltages. This was seen to be the case for mercury vapour as shown in Paper I. At lower velocities in hydrogen and helium, the effect of exchange must be included in the theory to explain the observed angular distributions. This may be a feature peculiar to these light elements, but it is possible that the effect of exchange should also be included in the calculation of the large angle scattering by heavy atoms at low velocities. It is also possible that the effect of exchange may become apparent in a careful comparison of accurate theoretical and experimental results for the small angle scattering by heavier atoms at moderate velocities.

In conclusion, we should like to express our thanks to Lord Rutherford and Dr. P. M. S. Blackett for their interest in the work, and to Dr. H. S. W. Massey for many interesting discussions.

Summary.

With an apparatus described in a previous paper the angular distributions of scattered electrons have been measured to 160° for 30–150-volt electrons. The elastic scattering has been investigated in hydrogen, carbon dioxide, nitrogen, neon, methane, hydrogen sulphide, and phosphine, while the inelastic

* 'Z. Physik,' vol. 70, p. 567 (1931).

† Henneberg, 'Naturwiss.,' vol. 30, p. 561 (1930); Massey and Mohr, 'Nature,' vol. 130, p. 276 (1932).

scattering has also been measured for the first four gases. The diffraction maxima and minima which appear in the inelastic angular distributions are closely similar to those obtained for the elastic scattering. The curves for the elastic scattering show a gradual change in form for successive elements in the periodic table.

With the aid of a simple picture involving merely the "size" of the atom, the general shape of the experimental curves for the elastic and inelastic scattering is qualitatively accounted for.



The Physical Principles of the Quantum Theory.

By G. TEMPLE, Ph.D., D.Sc., Imperial College of Science and Technology.

(Communicated by S. Chapman, F.R.S.—Received June 25, 1932.)

The object of this paper is to reformulate the principles of the quantum theory as a sequence of propositions which shall be either summary statements of standard experimental procedure or hypotheses concerning the results of experiment and having an immediate physical interpretation. It is shown that the standard process in micro-physics is a generalised spectral analysis, whose properties are simply expressible in symbolic form by means of projective or "idempotent" operators (Einzeloperatoren). It appears that only two hypotheses need be made and that these relate to the existence and properties of transition probabilities. From these fundamental principles, which have a direct physical significance, it is possible to deduce the subsidiary principles which form the accepted basis of the mathematical analysis of the quantum theory and which deal with the representation of quantum states and physical quantities by vectors and linear operators respectively.

In this paper the emphasis is laid on the experimental process determining a state of a system and on the associated operators rather than on the state itself or the vector representing it in the system space. Projective operators, which represent actual processes of measurement, and unitary operators, which represent actual transformations of systems of measurement, are given priority over the (statistical) operators which represent physical variables. This method of representation makes the physical meaning of the theory fundamental, instead of leaving it to be extracted from a purely mathematical system of non-commutative algebra or differential equations.

§ 1. *The General Principles of the Quantum Theory.*

The first step in the rational analysis of the principles of the quantum theory is the distinction between the "general principles," which are valid for any physical system, and the "special principles," which are characteristic of particular physical systems. The present paper is concerned only with the general principles. The special principles will be considered in a subsequent communication.

The fundamental concepts, in terms of which the general principles are formulated, are the concepts of a *system*, of the *states* of a system, of the *tran-*

sitions of a system from one state to another, and of the *probability* of these transitions. The general principles themselves relate exclusively to the mathematical representation of these physical concepts. They assert that a system is represented by a certain Hilbertian space \mathfrak{H} , the states of the system by rays or unit vectors, α, β, \dots , in the system space \mathfrak{H} , and the transitions of the system by unitary transformations, U, V, \dots , in \mathfrak{H} . The probability of a transition between two states is taken to be equal to $|(\alpha, \beta)|^2$, the squared modulus of the scalar product of the two rays α, β representing the states.

In this scheme a physical quantity or a dynamical variable is a derivative concept. It is associated with a group of unitary transformations, $\{U(s)\}$, depending upon one parameter s , and it is represented by the infinitesimal operator of this group,

$$P = \lim_{s \rightarrow 0} \partial U(s) / \partial s,$$

more precisely, by the operator $(\hbar/2\pi i) P$ where \hbar is subsequently identified with Planck's constant. The effective or average value of a physical quantity represented by an operator X in a state represented by a unit vector α is taken to be equal to $(X\alpha, \alpha)$, the scalar product of $X\alpha$ by α .

The two modes of exposition of the quantum theory, represented by the books of Dirac and Weyl, emphasise respectively the representation of states by vectors and the representation of quantities by operators. When the exposition is restricted to that part of the quantum theory which depends exclusively upon the general principles, Weyl's method is inevitable and it is this method—the representation of operators by algebras, groups and matrices—which is mainly employed here.

Both modes of exposition are purely deductive systems in which only the remote conclusions can be physically interpreted and experimentally proved. The principles themselves, as is clear, from the summary above, have no immediate physical significance; the object of this investigation is to analyse them into simple assumptions free from this defect. Until this problem is solved the quantum theory will present the anomaly of a physical theory incapable of giving a physical interpretation of its own principles.

§ 2. *Selective Operators.*

The general interpretation of the quantum theory given by Bohr, Heisenberg and Dirac has stressed the uncertain and unpredictable character of experimental observations. It is clearly as impossible to base the quantum theory on these *negative* qualities alone as to ground the theory of relativity

solely on the negative results of the Michelson-Morley experiment. On the other hand, a precise analysis of the *positive* characteristics of micro-physical measurement does furnish an entirely adequate physical basis for the quantum theory. It is argued here that the main experimental process of microscopic physics is a generalised form of spectral analysis. The essence of this type of process can be expressed quite simply in a symbolic form (§ 3), from which the accepted "general principles" of the quantum theory can be rigorously deduced with the aid of two auxiliary assumptions which are wholly physical in content.

Any physical experiment is an interaction between the system observed and the apparatus of observation. The action of the system on the apparatus in producing an observable record has its counterpart in the reaction of the apparatus on the system in producing an unobserved change of state. Hence only two types of experiment can yield unambiguous results—that type in which the action of the system on the apparatus is completely determined by the *initial* state of the system and that in which it is determined by the *final* state of the system. The characters of the system which can be inferred in these two cases are distinguished by Eddington† as "retrospective" and "contemporaneous" respectively. In an historical study of individual systems only the second type of experiment is efficacious.

In this special type of experiment the inference regarding the character of the observed system is valid precisely because the experiment *impresses* this character upon the system. Hence, if the same system is immediately subjected again to the same process, it will suffer no further change. If the experimental process is represented by the operator P this property of the process is symbolised by the equation

$$P^2 \equiv P \cdot P = P.$$

When the system is subjected to some other process, represented by the operator Q , the character impressed by Q may be incompatible with the character impressed by P , in which case the process Q will cause a transition of the system from one state to another. Symbolically, the compatibility or incompatibility of the characters impressed by P and Q is represented by the equations

$$PQ = QP \quad \text{or} \quad PQ \neq QP$$

respectively.

† "The Decline of Determinism," Presidential Address to the Mathematical Association, January 4, 1932.

Furthermore, two characters may be exclusive in the sense that no system possessing either character can receive the impress of the other character, *i.e.*, the process yields a null result. In symbols, exclusiveness implies that

$$PQ = O = QP,$$

where O is the null operator.

The preceding concepts have analogies in genetic biology. Here the process which impresses a character is selective generation or breeding. The propagation of a "pure line" which "breeds true" is analogous to the reproduction of the same character by repeated similar processes. In a pure line the determining character may be dominant or recessive, but these two kinds of characters are mutually exclusive, *i.e.*, recessive offspring cannot be bred from dominant parents in the pure line and *vice versa*. The incompatibility of dominant and recessive characters in the "mixed line" is illustrated by the generation of pure recessives from hybrid dominants.

In view of this analogy it is convenient to describe a process and its representative operator as "selective" if $P^2 = P$. (This avoids the awkward adjective "idempotent," introduced by Sylvester, and the untranslatable term "Einzeloperator" due to J. v. Neumann.) The state of a system is specified by the set of selective processes which produce no change in the system. These processes correspond to the totality of (compatible) characters possessed by the system in some state, and they completely describe the observable properties of that state.

§ 3. Spectral Sets of Operators.

The principles of the preceding section are of wide application. They are recognised in biology, and appear to be applicable to the psychology of conditioned reflexes. In physics, however, these general principles require specialisation in view of the quantitative nature of physical characters. Moreover, the theory has to be framed to include physical quantities which may vary either continuously or discretely. Hence the typical physical character is taken to be either that some variable ξ (such as a positional co-ordinate) has a numerical value not exceeding some prescribed number x , or that it has a value greater than x . The corresponding selective operators are written S_x or S_x' . We have to consider the properties of the sets of selective operators $\{S_x\}$, $\{S_x'\}$ where x has all possible values.

These properties can only be known by abstraction from the concrete selective processes actually employed in micro-physics. Now the type of

process which is of primary importance in this domain is exemplified in Stern and Gerlach's analysis of metallic vapours by an inhomogeneous magnetic field, in the analysis of positive rays by Aston's mass-spectrograph, and in the magnetic analysis of β -rays. All these methods present analogies with the spectral analysis of radiation. Their essential characteristic is the resolution of an inhomogeneous aggregate into (relatively) homogeneous parts.

In our symbolism the process of separation of a partial relatively homogeneous aggregate in which $x < \xi \leq y$ can be represented only by the operator

$$S_x'S_y = S_yS_x'.$$

Of course, such an aggregate may contain no members, in which case the operator $S_x'S_y$ is equivalent to the null operator O and we say that the region $x < \xi \leq y$ is absent from the "spectrum" of the variable for the particular aggregate subjected to analysis. The spectrum itself is then defined negatively as the set of values not excluded as "absent."

In this paper the term spectral analysis will be applied not only to the type of process illustrated above but will also be given a still wider significance. In the processes just cited, the aggregate to be analysed is composed of systems which are *simultaneously* passed through the analysing field but it is clear that no essential feature of the method of analysis would be varied if the individual systems were *separately* and successively subjected to the analytical process. Under these circumstances the complete process would consist of a multitude of separate experiments upon individual systems. The aggregate of these systems would then cease to be an actual collection and would become a mental fiction similar to a Gibbsian "ensemble."

From consideration of particular examples we can see that the necessary and sufficient conditions that two sets of complementary selective operators, $\{S_x\}$ and $\{S_x'\}$ should represent a process of spectral analysis are expressed by the following equations:—

$$\left. \begin{aligned} S_xS_y &= S_yS_x = S_x, & \text{if } x \leq y; \\ S_x'S_y &= S_yS_x', \\ S_y'S_x &= O = S_xS_y' \end{aligned} \right\} \text{if } x \leq y;$$

$$S_a = I, \quad S_b = O,$$

$$S_x + S_x' = I,$$

if a, b are the upper and lower bounds of the variable ξ and I is the identical

operator. Two complementary sets of spectral operators are called "spectral sets"† if they satisfy these conditions.

The mathematical analysis is considerably simplified if the continuous set of operators $\{S_\xi\}$ is replaced by the finite spectral set $\{P_n\}$ defined by $P_n = S_{x_n}' S_{x_{n+1}}$, where $\{x_n\}$ is a finite set of values of the variable ξ , dividing the complete domain of ξ into intervals of equal content, $x_{n+1} - x_n$, (or *a priori* probability‡); and if the character impressed by P_n is taken to be that ξ has the value x_n . The approximation involved in this substitution can be indefinitely sharpened by a subsequent passage to the limit in which the upper bound of the length of the intervals $x_{n+1} - x_n$ is made to tend to zero. The properties of the operators P_n are easily deduced from those of the set $\{S_n\}$. They are

$$P_n^2 = P_n,$$

$$P_m P_n = 0, \quad \text{if } m \neq n.$$

There is an obvious analogy between the set of operators $\{P_n\}$ and the geometrical operators which project a vector on to a set of orthogonal axes in multi-dimensional space, and this analogy is the guide to the subsequent theory of the representation of these operators. In anticipation of the results of this theory the operators will be described as "projective," and two projective operators P, Q such that $PQ = 0 = QP$, will be described as "orthogonal."

This analogy also suggests that the transition probability from the state determined by a projective operator P_m to the state determined by a projective operator Q_n should be the square of the cosine of the angle between the axes of P_m and Q_n , i.e., the characteristic§ (Spur) of the matrix representing the product $P_m Q_n$, since these quantities satisfy the obvious requirement that

$$\sum_n \text{Sp}(P_m Q_n) = 1.$$

Futhermore it suggests that the process represented by a numerical operator λ should be a change in the number of systems in the aggregate affected in the ratio $\lambda^2 : 1$. We shall adopt this last suggestion as a pure convention, obviously

† This term is due to Reisz. v. Neumann's expression "die Zerlegung der Einheit," which is translated by M. H. Stone as "the canonical resolution of the identity."

‡ Two intervals of a variable have equal *a priori* probability when they can be transformed into one another by a congruent transformation.

§ The characteristic of a matrix is the numerical sum of the elements in its principal diagonal.

not inconsistent with any physical fact, or the agreed properties of selective operators.

§ 4. Transition Probabilities.

The preceding section outlines the "descriptive" theory of spectral analysis, the axioms of which appear to be an immediate induction from the customary methods of physical measurement. The "metrical" theory of spectral analysis depends upon two additional assumptions regarding the probabilities of transition from one state to another. These assumptions are as follows:—

- (a) There is a definite probability p that a system in the state defined by a projective operator A ($= P_m$, say) will pass over into the state defined by a projective operator B ($= Q_n$, say) as a result of the process of spectral analysis defined by the spectral set, $\{Q_n\}$ to which B belongs.
- (b) This probability is completely determined by the operators A and B , and is the same for the transitions $B \rightarrow A$ and $A \rightarrow B$.

It follows from these assumptions that the two "partial" aggregates which are separated from any given aggregate by the processes represented by A and ABA will consist of systems in the same state, i.e., the state determined by A , and will differ only in the number of systems which they contain. The convention introduced at the end of the last section allows this result to be expressed in the form

$$ABA = \lambda A,$$

where λ is a numerical operator. Moreover, λ is evidently the probability p and is completely determined by the operator ABA , so that it may be rewritten in the form

$$\lambda = p(ABA).$$

Hence

$$ABA = p(ABA) \cdot A \quad \text{and} \quad p(ABA) = p(BAB).$$

The theory of the representation of projective operators is an immediate corollary to these two assumptions.† If $\{E_n\}$ is a complete set of orthogonal, projective operators, the matrix operator with character (E_j, E_n) of any projective operator P is defined to be

$$P_{jn} = E_j P E_n.$$

† A purely mathematical investigation is sufficient to deduce the possible irreducible representations of a *single* set of projective operators from the relations $P_n^2 = P_n$, $P_m P_n = 0$. The two physical assumptions made above are necessary to determine the representation of *two* distinct sets of projective operators $\{E_n\}$ and $\{P_n\}$. They ensure that these two sets have the *same* field of operation, i.e., that their representative matrices act upon the same set of vectors.

The two fundamental properties of matrix operators are :—

and
$$P_{ij}P_{kl} = E_i P E_j E_k P E_l = 0, \quad \text{if } j \neq k;$$

$$\begin{aligned} P_{ij}P_{jk} &= E_i P E_j P E_k \\ &= p (P E_j P) E_i P E_k \\ &= p (P_{jj}) P_{ik}. \end{aligned}$$

In particular the matrix operators of the operator E_n are all null operators except the one with character (E_n, E_n) for which

$$E_n,_{nn} = E_n.$$

§ 5. *The Matrix Representation of Operators.*

The fact that the matrix operators of P and of the set $\{E_n\}$, together with the numerical operators, form a field which is closed with respect to multiplication implies that these operators can be represented by matrices—in fact the operators $\{P_{ij}\}$ form the basis of a “complete matrix ring.”†

All the irreducible representations of the operators $\{P_{ij}\}$ are equivalent to a representation in which the only non-zero element in the matrix representing an operator of character (E_j, E_k) is in the row j and the column k . Also it is possible to choose the representation so that this element has the form $x_j^* x_k$ where $\{x_k\}$ is a set of complex numbers and $\{x_k^*\}$ their conjugates. It is clear that

$$x_k^* x_k = p (P_{kk}),$$

so that only the moduli of the numbers $\{x_k\}$ are determinate; their phases can be prescribed arbitrarily.

Moreover, it can be shown that the representations of the matrix operators of two distinct projective operators P and Q are simultaneously reducible to the canonical form defined above. For,

and
$$P_{ij}Q_{kl} = 0, \quad \text{if } j \neq k,$$

$$\begin{aligned} P_{ij}Q_{jk} &= E_i \cdot P E_j Q \cdot E_k \\ &= \text{a matrix operator of character } (E_i, E_k), \end{aligned}$$

hence the result follows at once.

It remains to consider the representation of the projective operators themselves. Although the general concept of the *sum* of two or more operators

† Van der Waerden, “*Moderne Algebra*” (1931).

cannot yet be defined in physical language (see § 9) there is an obvious sense in which a projective operator P is the *sum* of its matrix operators $\{P_{ij}\}$, *i.e.*, if ϕ denotes any aggregate of systems, the aggregate of systems $P\phi$ is the sum of the aggregates $P_{ij}\phi$. This result follows from the fact that the system of projective operators $\{E_j\}$ is "complete," *i.e.*, any aggregate is exhaustively analysed by the corresponding process of spectral analysis. This is the physical significance of the "double Peirce reduction" of P ,

$$P = \sum_{j, k} P_{jk}.$$

This relation shows that a projective operator P can be represented by a matrix which is the sum of the matrices representing its matrix operators P_{jk} .

Again, if A and B are two projective operators the definition of the *sum* $\sum_j A E_j B$ as equal to AB can be similarly justified. Hence, if $C = AB$,

$$\begin{aligned} C_{jk} &= E_j A B E_k = \sum_l E_j A E_l \cdot E_l B E_k \\ &= \sum_l A_{jl} B_{lk}. \end{aligned}$$

Therefore the matrix representing AB is the ordinary matrix product of the matrices representing A and B .

The representation of the operators by matrices implies a simultaneous representation of the states of a system by vectors in the space \mathfrak{H} . The action of the physical processes denoted by the operators upon the states of the system is then represented by the transformation of the vectors by the matrices. The state which is determined by the projective operator P_n is represented by the vector σ_n , which is invariant under the transformation by the matrix representing P_n . It is convenient to use the same symbols to denote the physical process, the corresponding operator and its representative matrix; also to use the same symbols for the state and its representative vector. With this convention

$$\begin{aligned} P_n \sigma_n &= \sigma_n, \\ P_m \sigma_n &= 0, \quad \text{if } m \neq n. \end{aligned}$$

With the convention that the states determined by the projective operators A and B are represented by vectors α and β , of unit length, it follows that the probability of the transitions $\alpha \rightarrow \beta$ or $\beta \rightarrow \alpha$ is $|(\alpha, \beta)|^2$.

§ 6. Average Values of Individual Variables.

It is now possible to deduce the result anticipated at the end of § 3, *i.e.*, that the probability of the transitions $A \rightarrow B$ or $B \rightarrow A$ equals the characteristic of

the matrix representing the operator ABA . The diagonal matrix elements of A are the numbers $p(A_{kk})$, i.e., the numbers $p(E_k A E_k)$ which are the probabilities of the transition $A \rightarrow E_k$. Since the set of operators $\{E_k\}$ is complete, the sum of these numbers is unity, i.e.,

$$\text{Sp} A = 1.$$

Now

$$ABA = p(ABA) \cdot A,$$

whence

$$\begin{aligned} \text{Sp}(ABA) &= p(ABA) \cdot \text{Sp} A \\ &= p(ABA), \end{aligned}$$

—the required result. This expression for the transition probability is exact, whereas the first expression deduced below for the average value of a variable is only an approximation, subsequently made exact by a passage to a limit.

We can now obtain a simple expression for the average value of a variable ξ in the state specified by the projective operator A . An average value can be defined only in terms of the experimental process by which it is actually determined. To determine the average value it is necessary to analyse an aggregate of systems in the specified state by means of the set of operators $\{P_n\} = \{S'_n S_{n+1}\}$. The probability of the transition $A \rightarrow P_n$ is simply $\text{Sp}(AP_n)$. Hence the average value of ξ is approximately $\sum_n x_n \text{Sp}(AP_n)$.

If we introduce the matrix S defined by the equation,

$$S = \sum_n x_n P_n,$$

the average value of the variable ξ in the state determined by A is concisely expressible as

$$E(\xi) = \text{Sp}(AS).^\dagger$$

This approximation to the average value of ξ can be sharpened by increasing the number of operators in the set $\{P_j\}$. The exact value obtained on proceeding to the limit is given by the Stieltjes' integral

$$\int_{-\infty}^{+\infty} x d\omega(x),$$

where

$$\omega(x) = \text{Sp}(AS_x).$$

The main problem of quantum theory is the determination of average values,

[†] $E(\)$ is the symbol used by v. Neumann for the "Erwartungswert."

and for this problem the variable ξ is represented by the operator $S = \sum_n x_n P_n$, or, accurately by

$$\int_{-\infty}^{+\infty} x dS_x.$$

Hence S may be called the statistical operator of ξ . Conversely, any complete set of projective operators $\{Q_n\}$ together with a set of numbers $\{y_n\}$ determines a statistical operator, $T = \sum_n y_n Q_n$ and thus represents some variable η .

Finally, we note that the average value of $f(\xi)$, where f is any polynomial function of ξ or the limit of a sequence of such function, is given by

$$E(f) = \sum_n f(x_n) \text{Sp}(AP_n),$$

i.e., by

$$\text{Sp}(Af(S)),$$

where

$$f(S) = \sum_n f(x_n) P_n.$$

§ 7. Congruent Transformations and Unitary Operators.

Further progress in the theory of the representation of operators requires the powerful methods of group theory introduced by Weyl.[†] The argument is that a group of congruent transformations of a physical quantity ξ induces a corresponding group of unitary transformations in the system space \mathfrak{H} and that the infinitesimal operator of this group corresponds to the variable dynamically conjugate to ξ . The sum of two operators can then be defined in terms of the product of the finite operators which they generate when regarded as infinitesimal operators.

Congruent transformations arise from the comparison of different methods of determining the physical characteristics of the same system. Hence a congruent transformation of a variable is simply a permutation of the proper states of this variable, i.e., the states determined by the projective operators of ξ . Hence the corresponding transformation U in the system space \mathfrak{H} ,

$$\sigma_n \rightarrow U\sigma_n = \sigma'_n,$$

is unitary, so that

$$(U\sigma_n, U\sigma_n) = 1.$$

The matrix operators and matrix elements of a unitary operator U are defined as in the case of projective operators. If $\{E_n\}$ is the spectral set of

[†] "The Theory of Groups and Quantum Mechanics" (Eng. trans., 1931), p. 185.

projective operators taken as the basis, the matrix operator of character (E_j, E_k) is

$$U_{jk} = E_j U E_k.$$

From the unitary property of U it follows, by the usual argument, that the j , k -matrix element of U and the k , j -matrix element of its inverse U^{-1} are conjugate complex numbers.

When the variable which is the subject of a congruent transformation is the time, the corresponding unitary operator U transforms the state of a system at any given time into its state at some subsequent time and thus determines the historical development of the system. If the initial state of the system is a proper state, α_j , of some projective operator, A_j , belonging to a spectral set $\{A_j\}$, the final state, β_j , of the system is not necessarily a proper state of any of these operators. Nevertheless the final state is completely determinate, and its properties may be specified by subjecting the systems in this state to the process of spectral analysis represented by the set of operators $\{A_j\}$. The double transition, $\alpha_j \rightarrow \beta_j \rightarrow A_k \beta_j$, is determined by the operator $A_k U$. Hence the probability of the transition from a proper state of A_j to a proper state of A_k under the influence of U will be zero if

$$A_k U \alpha_j = 0,$$

i.e., if

$$U_{kj} = A_k U A_j = 0.$$

Thus the evanescent matrix elements of U with basis $\{A_j\}$ determine what transitions the operator U cannot produce.

In general, the probability of the transition $\alpha_j \rightarrow A_k \beta_j$ is determined by A_k and B_j —the projective operator specifying the state β_j . In the notation of § 4 this transition probability is $p(A_k B_j A_k)$.

Now the transformation

$$\alpha_j \rightarrow U \alpha_j = \beta_j,$$

implies the transformation

$$A_j \rightarrow U A_j U^{-1} = B_j.$$

Hence

$$\begin{aligned} p(A_k B_j A_k) &= p(A_k U A_j U^{-1} A_k) \\ &= p(A_k U A_j \cdot A_j U^{-1} A_k) \\ &= p(U_{kj} \cdot U_{jk}^{-1}) = |u_{kj}|^2, \end{aligned}$$

where u_{kj} is the only surviving matrix element of U_{kj} .

§ 8. Conjugate Variables.

The set of congruent transformations of a single variable ξ form a group, and, since there is isomorphic correspondence between these transformations and the unitary transformations which they induce in the system space, these latter must also form a group. An individual transformation of this group, $V(s)$ will be distinguished by a parameter s which is to be chosen (as is always possible) so that

$$V(s) V(t) = V(s + t).$$

It will be shown in this section that the group of unitary operators $\{V(s)\}$ represents the variable dynamically conjugate to ξ .

Let Q be the projective operator determining a state ψ which is unchanged by the group of unitary operators $\{V(s)\}$. Then

$$Q = V^{-1}(s) Q V(s),$$

for all s .

Let P_j, P_k be two projective operators determining states σ_j, σ_k which are transformed into one another by the unitary operators $V(t), V^{-1}(t)$. Then

$$P_j = V^{-1}(t) P_k V(t).$$

Hence

$$\begin{aligned} \text{Sp}(P_j Q) &= \text{Sp} V^{-1}(t) P_k V(t) \cdot V^{-1}(t) Q V(t) \\ &= \text{Sp} V^{-1}(t) \cdot P_k Q \cdot V(t) \\ &= \text{Sp}(P_k Q). \end{aligned}$$

Accordingly there is the same probability of transition from the state determined by Q to any proper state of the variable ξ . Also, since $\sum_j \text{Sp}(P_j Q) = 1$, these transition probabilities all have the same value, N^{-1} , where N is the number of projective operators in the complete set $\{P_j\}$.†

Accordingly if $\{Q_k\}$ is the set of projective operators which are invariant under the group $\{V(s)\}$, then all the transition probabilities $P_j \longleftrightarrow Q_k$ are equal. Under these circumstances we say that the two sets of operators and the corresponding sets of states, are *conjugate*.

It follows from standard matrix theory that the totality of projective operators $\{Q_k\}$ which are invariant under $\{V(s)\}$ form a complete spectral set, and that if those operators are taken as a basis, the unitary operator $V(s)$ is represented by a diagonal matrix, in which the diagonal element v_{kk} has the

† This assumes that the domain of ξ is finite, i.e., ξ is a cyclic variable. Non-cyclic variables (if such exist) must be treated as limiting cases.

form $\exp (isy_k)$ where y_k is a real number. Hence, for matrices, we have the result,

$$V(s) = \sum_k \exp (isy_k) \cdot Q_k.$$

To interpret this matricial equation, we note that, in accordance with the result of § 6 the set of real numbers $\{y_k\}$ and the spectral set $\{Q_k\}$ are associated with some variable η which is represented by the statistical operator

$$T = \sum_k y_k Q_k,$$

whence

$$V(s) = \exp (isT).$$

Thus the same variable η is represented by the statistical operator Q in the theory of average values and by the unitary operators $\{V(s)\}$ in the theory of congruent transformations.

The concept of two dynamically conjugate variables cannot be bodily transferred from classical theory to quantum theory, as an exact determination of the simultaneous values of a co-ordinate and its momentum is experimentally impossible. But the rôle of the variable conjugate to ξ is played in the quantum theory by the variable η defined indirectly above. This identification of the nature of η is strengthened by the following analogy:—in classical theory a definite constant value of a momentum implies that it is equally probable that the value of the co-ordinate lies in any two intervals of equal length. Similarly in quantum theory, for a system in the state defined by Q_k (for which η has the definite value y_k) there is equal probability for all the transitions $Q_k \rightarrow P_j$, ($j = 1, 2, \dots$).

§ 9. *Average Values of the Sum or Difference of Two Variables.*

In quantum theory the concept of average values is subject to severe limitations, which are imposed by the very nature of the experiments which have to be made in order to measure them. Thus, if ξ and η are any two variables, although we can experimentally define the average values of ξ , of η and of $\xi + \eta$ for any prescribed state, we cannot define the average value of $\xi\eta$ unless the statistical operators S and T commute, and even the relation

$$E(\xi) + E(\eta) = E(\xi + \eta)$$

is not self-evidently true.

To establish this relation for variables which are conjugate to co-ordinates susceptible of *continuous* variation it is sufficient to utilise the double repre-

sensation of such variables as statistical operators and as unitary operators. The variable ξ is represented either by the statistical operator

$$S = \sum x_n P_n,$$

or by the group of unitary operators

$$U(\lambda) = \exp(i\lambda S).$$

It is clear that if the variable conjugate to ξ is continuously variable, the group $\{U(\lambda)\}$ is also continuous and that its infinitesimal operator is

$$[\partial U(\lambda)/\partial \lambda]_{\lambda=0} = iS.$$

Now let η be represented by

$$T = \sum y_n Q_n,$$

and by the group

$$V(\lambda) = \exp(i\lambda T).$$

Although it is not immediately evident what *statistical* operator will represent the sum $\xi + \eta$ it is clear that the representative *group* will be $\{U(\lambda) V(\lambda)\}$ for this represents the joint effect of $V(\lambda)$ followed by $U(\lambda)$. Hence we can deduce that the statistical operator of $\xi + \eta$ is the infinitesimal operator of the group,

$$-i\{U(\lambda) V(\lambda)\},$$

i.e.,

$$S + T.$$

Hence, if S and T represent the statistical operators of the variables ξ and η , then $S + T$ represents $\xi + \eta$. Similarly $\xi - \eta$ is represented by the group $\{U(\lambda) \cdot V^{-1}(\lambda)\}$ and by the statistical operator $S - T$, and, in general, the statistical operator of $a\xi + b\eta$ is $aS + bT$, where a, b are any two numbers.

This result completes the theory of the representation of physical quantities by operators. From this point the "general" quantum theory can be developed as in Weyl's treatise. The exchange relations for conjugate variables follow from their representation by unitary operators and the theory of the angular momentum and polarisation operators is deducible from the isomorphism of the congruence groups and their associated unitary groups in system space. The problems of "special" quantum theory, and, in particular, the theory of the Hamiltonian operator will be discussed in a future paper.

Summary.

An analysis of the general nature of physical measurement shows that quantum states are defined by certain "selective" processes, represented

by operators which satisfy the condition, $P^2 = P$. The general principles of the representation of quantum states and physical variables by sectors and matrices in Hilbertian space can then be deduced from two physical hypotheses:—

(a) that (under specified conditions) there is a definite probability for the transition from the state represented by P to the state represented by Q .

(b) that this probability is the same for the transitions $P \rightarrow Q$ and $Q \rightarrow P$.

The association of congruent transformations with unitary operators leads to the theory of dynamically conjugate variables, and completes the theory of the "effective" value of a variable, averaged over all systems in a definite state.

A Unique Electrode Potential Characteristic of a Metal, and a Theory for the Mechanism of Electrode Potential.

By A. L. MCAULAY and E. C. R. SPOONER.

Physics Laboratory, University of Tasmania.

(Communicated by T. M. Lowry, F.R.S.—Received June 29, 1932.)

1. *General.*

Metal electrodes immersed in air free solutions of KCl were found to come to a definite reproducible potential. From the Nernst expression, this potential corresponds to a certain concentration in the solution of the ion of the electrode metal. Chemical analysis showed that the concentration that actually exists in the bulk of the solution is much smaller than that deduced from the potential, even after electrode and solution have been in contact for weeks.

It was felt that a knowledge of the cause of this definite potential was fundamental to the general investigation of the potentials of metals in solutions of varying aeration, p_H , and salt concentration, a subject of great practical importance which is at present in rather an indefinite and unsatisfactory condition.

Work was undertaken with a view to obtaining information about the potentials of metals immersed in air-free solutions containing few of their own ions. In most of the experiments cadmium was used as an electrode, but check experiments have been made using other metals.

The results disclose a single potential to which a metal tends to come, and leads to a theory for the mechanism which produces this. Further work was done on solutions which had been aerated.

2. Experimental.

The apparatus and technique were as described in a paper by McAulay and Bastow,* except that oxygen-free hydrogen replaced the atmosphere of nitrogen over the solutions, and connection between flask and calomel half-cell was made by means of an agar bridge of the same salt concentration as in the bulk solution.

3. A Potential Characteristic of a Metal and Water Alone.

For concentrations less than a certain sharp minimum the potential of cadmium in a solution is unaffected by the concentration of its own ions, in contradiction to a direct application of Nernst's theory.

Provided the solution is air-free and reasonably dilute, the potential it then attains is characteristic of the electrode alone and independent of the nature of the solution. The only constituents of the system which have not been varied while the potential remained constant are water and the electrode metal; and the mechanism of the production of the potential must therefore be constructed from these alone.

Potentials of a cadmium electrode in various rigidly air-free solutions, Table I, illustrate this point.

Table I.

Solution.		Potential at 18° C. against sat. cal. electrode.	Solution.		Potential at 18° C. against sat. cal. electrode.
1	M/500 K ₂ SO ₄	0.788	9	M/500 KCl	0.784
2	M/500 K ₂ SO ₄	0.787		M/2 × 10 ⁴ CdCl ₂	
	M/10 ⁵ CdSO ₄		10	M/10 ⁷ CdCl ₂	0.792
3	M/10 ⁵ CdSO ₄	0.787	11	M/10 ⁶ CdCl ₂	0.786
4	M/500 KCl	0.782	12	M/10 ⁵ CdCl ₂	0.786
5	M/500 KCl	0.786	13	M/10 ⁵ HCl	0.787
	M/10 ⁷ CdCl ₂		14	M/10 ⁴ HCl	0.782
6	M/500 KCl	0.786	15	M/10 ³ HCl	0.784
	M/2.5 × 10 ⁵ CdCl ₂		16	M/10 ⁵ NaOH	0.780
7	M/500 KCl	0.793	17	M/10 ⁴ NaOH	0.795
	M/5 × 10 ⁵ CdCl ₂		18	M/10 ⁴ H ₂ SO ₄	0.790
8	M/500 KCl	0.787			
	M/10 ⁵ CdCl ₂				

Mean potential for all solutions, 0.787.

* McAulay and Bastow, 'J. Chem. Soc.,' p. 85 (1929).

It will be seen that the greatest variation from the mean value for the 18 different solutions considered is 8 millivolts, and that no system exists among the variations. A direct application of Nernst's theory would predict a difference of nearly 80 millivolts between 9 and 10 or 9 and 5.

4. Theory of the Mechanism which produces Electrode Potential.

The usual Nernst theory of electrode potential may be stated in a kinetic form as follows. There is an exchange of metal ions between solution and electrode, a postulated "solution pressure" forcing a stream of ions out of the metal, and their osmotic pressure in the solution urging a stream on to it. The electrode takes up a potential such that the supply of ions due to their migration under the electric field is equal and opposite to the difference between the two streams. The following extension is made to the theory to explain the above facts, and further experimental results are examined by its aid.

It is evident from a consideration of the results given in the last section that a theory of the mechanism of the production of electrode potential must involve primarily an interaction between metal and water only, the concentration of ions in the solution appearing as a secondary factor. The removal of the positive ions from the metal is believed to be effected by the forces that act between them and the negative ends of the polar water molecules. The heat motion of the water molecules brings them in contact with the ions, as a result there is an accumulation of metal ions in the immediate neighbourhood of the electrode, and equilibrium is established when the streams of ions due to this cause, to osmotic pressure and to migration under the field set up by the escaping ions, balance.

The ions will be carried a mean distance, t , by the water molecules and will migrate back to the electrode. The equilibrium will be practically unaffected by the concentration of the ions of the electrode metal in the bulk solution, until this bulk concentration reaches a critical value about equal to that due to the ions escaping from the electrode. This is in accordance with observation.

5. Concentration of Metal Ions in Layer.

The concentration of ions in the layer surrounding the electrode can be arrived at in two completely independent ways, one of which gives the total concentration, and the other the concentration of uncombined ions. The total concentration is obtained by finding what concentration of a salt of the electrode metal must be added to a solution before the potential changes. This potential

change takes place sharply, the potential at higher concentrations following the Nernst law. The critical concentration for cadmium is about 4×10^{-5} M and is reproducible under widely varying experimental conditions.

The concentration of uncombined ions is given directly by the potential of the electrode in ion-free bulk solution, as 2.5×10^{-5} M, which is in good agreement with the value for total concentration.

6. *Mean Distance from Electrode to which Ions are carried by Water Molecules.*

The theory can be checked immediately by an approximate calculation of the thickness of the layer of ions surrounding the metal, or, what amounts to the same thing, the mean distance, t , to which they are carried by the water molecules. The approximate assumption will be made that there is a layer of ions of uniform concentration, c gm. mols. per litre, and depth, t , in the neighbourhood of the electrode, and that its capacity is given by that of a parallel plate condenser of thickness $t/2$.

The following equations then give t for cadmium in terms of the electrode potential

$$V - V_0 = 0.0288 \log_{10} (\delta c)$$

and

$$\frac{ct \times 1.59 \times 10^{-19} \times 2}{1.66 \times 10^{-24} \times 1000} = V \times \frac{10^{-12}}{4\pi t/2},$$

where V is the absolute electrode potential.

The value obtained for t is not greatly affected by the value assumed for the absolute zero of electrode potential. If the usual figure -0.53 on the saturated calomel scale is taken, t is given as about 3×10^{-6} cm., while if a figure as positive as $+0.23$ is assumed, which recent work suggests may be correct* t is given as about 6×10^{-6} . Both of these are within the range of reasonable values, as may be seen from the following considerations. t must be fairly large compared with the diameter of a molecule, and would for this reason have a lower limit of the order of 10^{-6} , while the observed fact that violent stirring does not change the potential in rigidly air-free solutions would set an upper limit of the order of 10^{-4} . The calculated values lie within this range.

7. *Effect of Depressing Ionic Concentration in Layer.*

The above theory of electrode potential may be tested by experiments in which measurements are made of electrode potentials with varying con-

* Jean Billiter, 'Trans. Amer. Electrochem. Soc.,' vol. 57, p. 351 (1930).

concentrations of the metallic ion in a set of air-free solutions having widely different concentrations of KCl.

Increasing the concentration of KCl in a solution containing CdCl_2 forces the Cd ions into molecular combination with Cl ions, and so changes the ratio of the concentrations of combined and uncombined ions. A series of curves in which potentials are plotted against the negative logarithm of the total (*i.e.*, uncombined + combined) Cd ion concentration in solutions of widely different concentration of KCl are shown in fig. 1. From them it is possible to deduce

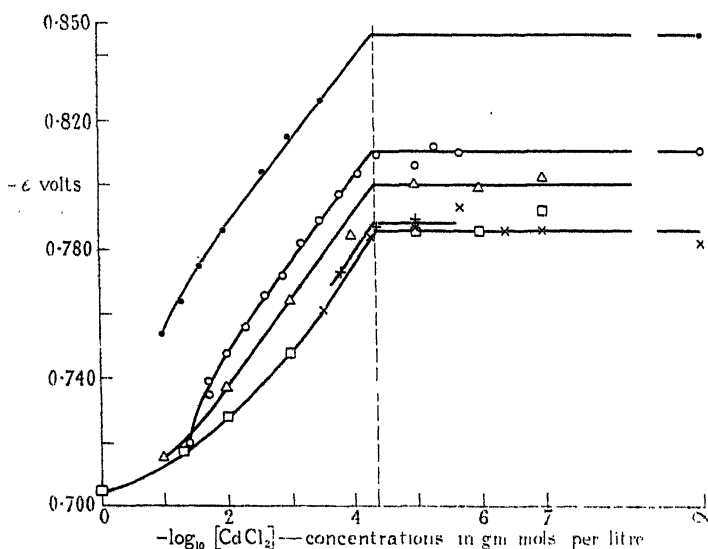


FIG. 1.—• M/1 KCl. ○ M/10 KCl. △ M/50 KCl. + M/100 KCl. × M/500 KCl.
□ CdCl_2 alone.

the concentration of uncombined Cd ions in the layer, and compare it with the total Cd concentration in the layer. It is the variation of these quantities with the concentration of the anion that is instructive.

The theory given above predicts that in the absence of Cd ions in the bulk solution

$$V = A + c$$

$$V - V_0 = 0.0288 \log_{10} kc,$$

where c is the total Cd ion concentration in the layer, V the absolute potential, kc is the ionic concentration and A is a constant. Thus, even if the most unfavourable value of the absolute zero of potential is taken, a depression of the ionic concentration a hundred-fold by the addition of KCl will only change the value of c by about 3 : 4.

This prediction is verified by the curves, which show a change of this order in ionic concentration (given by the potentials of the horizontal portions of the curves), while there is no sensible change in the total concentration in the layer (given by the abscissa at the point at which a curve changes from a sloping to a horizontal line).

8. The Effect of Exposure to Air.

If the cadmium electrode in its solution is exposed to air, its potential becomes unreproducible and takes up a much more positive value than it has in air-free solution.

Although in aerated solutions readings are numerically unreproducible, information can be obtained regarding the alteration of potential in the aerated state due to selected changes in condition.

Some experiments were made with the object of obtaining information concerning the change brought about in the layer of metal ions surrounding an electrode in air-free solution when air is allowed access to the electrode.

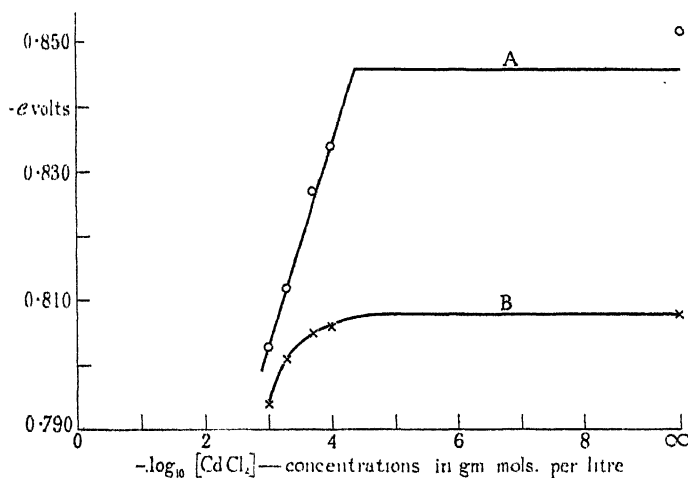


FIG. 2.—A reading taken after immersion of electrode for $\frac{1}{2}$ hour.

Preliminary experiments showed that it was better to use a solution fairly strong in KCl, as a more stable quasi-equilibrium could be obtained in this way in aerated conditions. They also showed that with a certain arbitrary standard method of aerating, the total Cd ion concentration in the layer was somewhere between 10^{-3} and 10^{-4} .

Fig. 2 shows potentials of Cd electrodes in M/1 KCl at different concentrations of CdCl_2 the abscissa giving the negative logarithms of these concentrations.

Curve A gives the potentials before air is admitted to the solutions, curve B the potentials afterwards.

It will be noticed at once that the potential when it has been made more positive by aeration is much less sensitive to change in concentration of Cd ions in the bulk solution, but when a sufficiently great concentration is reached the potential is determined by this factor as with air-free solutions. The significant fact is that the critical Cd ion concentration below which further decrease has little effect on the potential comes at a higher concentration than with air-free solutions. This must mean that the concentration of ions in the layer that surrounds the electrode is greater when an oxide is present than when it is not. The more positive potential produced by aeration must therefore in this case be due to a freer supply of ions by the oxide than by the metal, and not to partial protection of the surface by an impervious film.

9. *Summary.*

If certain limits are not exceeded, experiment shows that the potential of cadmium in an aqueous solution is independent of all changes in the character and constitution of the solution. It is therefore concluded that the electrode potential must have its origin in an interaction between metal and water only.

A theory is advanced for the origin of the electrode potential of a metal immersed in any aqueous solution, and experiments quoted in support. The theory is as follows. Water molecules, which are strongly polar, bombard owing to their heat motion the metal ions that exist in a metal. As a result of this process a stream of ions is drawn into the solution to a mean distance, t , from the electrode surface, and equilibrium is reached when their removal is balanced by their migration back under the field produced by their abstraction. This results in a layer of ions whose thickness is of the order of t surrounding the metal.

The following are consequences of the theory which are verified by experiment :—

- (i) The electrode potential must be independent of all change in the character and constitution of the solution, provided that the concentration of electrode metal ions in the bulk of the solution is less than that in the layer, that no kation is present whose solution pressure is low compared with that of the electrode metal, and that the anionic

concentration in the solution is not high enough to depress greatly the ionic concentration of the electrode metal in the layer.

- (ii) The thickness of the layer must lie between 10^{-6} and 10^{-4} cm.
- (iii) Large changes in the ionic concentration in the layer, produced by making large changes in the anionic concentration of the solution, and so forcing the ions of the layer into undissociated molecules, must leave nearly unchanged the total concentration (dissociated + undissociated) in the layer.

In the case of cadmium, experiment gives the thickness of the layer as about 5×10^{-5} cm., and the total concentration in it as about 4×10^{-5} gm. mol. per litre. The potential of Cd in air-free dilute solutions of any kind is 0.787 volt against the saturated calomel electrode (temperature 18° C.).

When an electrode is exposed to air, the potential becomes more positive and unreproducible. Evidence is given for the belief that the layer of positive ions which surround the electrode is more concentrated in this case, and that this and not the formation of an impervious film is the cause of the change of potential in the case of cadmium.

Triboelectricity and Friction. VII.—Quantitative Results for Metals and other Solid Elements, with Silica.

By Professor P. E. SHAW, M.A., D.Sc., and E. W. L. LEAVEY, M.Sc., University College, Nottingham.

(Communicated by Sir William Hardy, F.R.S.—Received July 4, 1932.)

For research in triboelectricity (*i.e.*, frictional electricity), metals and silica are pre-eminently suitable; they are hard, chemically simple, and their surfaces can easily be cleansed. Systematic work on these materials is described in a previous paper of this series.* We have gone over the ground again using improved apparatus and methods; and the results, which in the earlier paper were little better than qualitative, may now claim quantitative rank. Comparison of the two sets of results shows that they are in substantial qualitative agreement.

The process throughout this series of papers is to rub one solid on the other in a machine which maintains constant pressure and constant length of rub; in the present case a metal or other rod is rubbed on a silica tube. The silica is surrounded in part by a metal inductor connected by wire to an electrometer, so that the charge produced by friction on the silica induces a charge on the inductor and electrometer. A reading is taken on the electrometer; and by

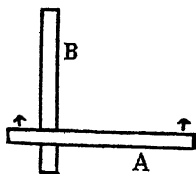


FIG. 1(a).

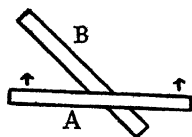


FIG. 1(b).

calibrating the system with known voltages, the electrometer readings are known in volts. Since the capacity of the charged system remains constant, the charges generated vary as the readings in volts. The former method of rubbing the surfaces is shown in fig. 1 (a) where the upper rod A in its stroke bears *with one spot* on a *succession of spots* of the lower rod B. Thus the single spot in the upper rod rubbed throughout the stroke becomes hotter and more

* Shaw and Jex, 'Proc. Roy. Soc.,' A, vol. 118, p. 97 (1928).

strained than any spot on the lower rod.* Such differential treatment of the two surfaces with its complicating effect on the physical action of the friction, we now avoid, as follows: place the lower rod at 45° to the line of run of the upper one, fig. 1 (b), so that when A moves forward, a succession of points on it come in contact with a succession of points in B. In this way the treatment of the two surfaces is equalised.

The diagonal method has another advantage. A particle of dust or other impurity caught between the rubbing rods may be carried along the whole stroke in the old method, producing false values of charge. In the diagonal method this trouble is eliminated.

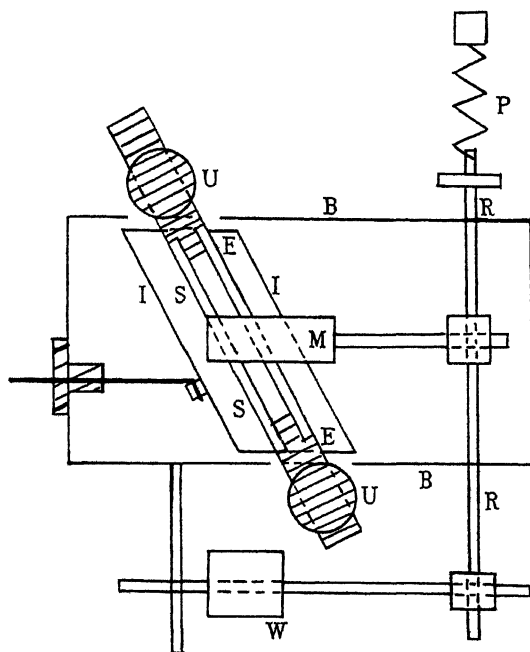


FIG. 2.—Plan of the rubbing apparatus. SS is the silica tube 6 inches in length; M is the metal rod.

The plan of the rubbing apparatus is shown in fig. 2. The stout-walled tube of silica, S, is held horizontal by ebonite plugs E, clamped in uprights U. The skew cylindrical inductor I, which partly surrounds the silica, is joined by wire to the electrometer, and is insulated by amber blocks, not shown. The metal rod M is attached to the rod R, which has two motions: (1) a movement along its length under the pull of spring P; (2) a rotation on its axis by which the rod

* Shaw and Jex, *loc. cit.* p. 99, (1927); Shaw and Hanstock, 'Proc. Roy. Soc.,' A, vol. 128, pp. 474, 480 (1930).

M can be lifted clear of the rod S. The length of rub is 30 mm. Length and pressure are constant throughout the experiments. Enclosing the two charging rods M and S is a brass box B. The sides of the box provide bearings for the rod R; it has large holes through which the silica rod system penetrates at each end as shown; the box is earthed and with its lid it provides a complete screen for the charges. The wire from the inductor in the apparatus (A)

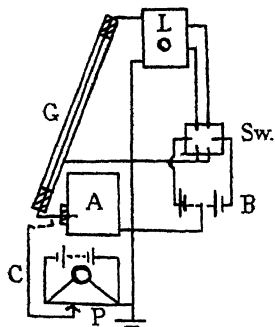


FIG. 3.—Scheme of the whole apparatus. A the rubbing box is joined to the electrometer L. B is the high tension battery with switch Sw. C is the calibrator with potentiometer P.

passes to the electrometer L, fig. 3, through an earthed tube G in the usual way. The calibrating system C is shown, and the high tension battery B of 120 volts with switch Sw is on the right.

The method of operating the apparatus differs in some ways from that described in previous papers. We now keep the discharging radium in a lead box away from the rubbing apparatus except when in use. The rods are rubbed together *only once*, and the electrometer is read; the radium is introduced into the apparatus, and, when the charges have been eliminated it is replaced in the lead box. For the next stroke the silica tube is turned slightly so as to present a new surface for the next rub, and the metal rod is well scraped again to renew the surface. In this way quite fresh surfaces are presented for each stroke. Some 20–30 strokes are taken for every pair of solids. We use a string electrometer which proves to be sensitive and reliable, and, being deadbeat, is far more convenient for the purpose than a quadrant electrometer.

Materials.—The vitreous silica tubes are from the Thermal Syndicate, Ltd., who give the analysis as 99·8 per cent. pure SiO_2 . The metals and metalloids used may be divided into two classes (a) those of special purity, which we have used in previous work,* viz., Au, Pt, Ag, Cu, Fe, Ni, Al, Sn, Cd, Sb, Cr, which are all better than 99·7 per cent. pure; (b) those of commercial purity, Pb, Zn, Bi, Co, Tl, Se. We use pure materials for preference but our experience is that, in this entirely surface work, body impurity counts little in comparison with surface contamination. This matter of surface variability has ever been the bugbear of workers in friction and frictional electricity. Solid surfaces are prone to oxidation and other chemical action, as well as adsorption of water

* Shaw and Jex, 'Proc. Roy. Soc.,' A, vol. 118, p. 98 (1928).

vapour or other vapour in the air. And it may well be that an ordinary, say, 1 per cent., impurity affects our results only, if at all, by its own special effect on oxidation and adsorption.

All surfaces are cleansed just before use, the silica by heating in a blowpipe for a minute or two, the other materials by first heating well in a flame and then scraping with a razor or file cleaned in the flame; a razor blade is preferable to a file as it leaves a smooth surface, but the file is necessary for hard substances such as Cr, Fe, Ni, Co. It should be observed that the scraping or filing leaves the surfaces strained, whereas for standard purposes they ought to be annealed. But as annealing involves heating and oxidation, we choose the lesser evil of strain. In any case the effect of such strain is slight, and, for the same material and method of scraping, is constant.

The surfaces thus prepared give charges on rubbing which repeat with as good consistency as can be expected in the uncertain rough-and-tumble of the sliding friction of solids.*

The uncertainty of the charges found in this subject is an old story; a solid surface is very sensitive to, and retentive of, slight chemical or physical influence, and the experimenter, however experienced, is liable at any time to stumble on fresh evidence of this fact. As an example, our practice is to cleanse platinum by roasting it in a blowpipe, and then gently to scrape the precious metal. On one occasion some cotton which had been exhaustively cleansed† was rubbed on the clean platinum; immediately the metal surface fell, like Lucifer, from its exalted place in the list, Table I, *i.e.*, from the head of the positive series, to a negative value. It is vain to enquire whether the change due to the cotton was physical or chemical. Probably it was chemical, but the lesson is obvious. In the long history of this subject, from the time of Gilbert, results have ever proved difficult of repetition (or even "fortuitous") so that hitherto nothing better than a qualitative list has been attempted. But such a list is unsatisfactory; it represents merely a gradual descent by steps from positive to negative values, but it does not indicate the depth of the successive steps. It may be that a particular step in values at one place on the scale may be as great as that embracing ten consecutive steps elsewhere. Such disparity in the steps can only be revealed by a quantitative list.

Encouraged by the improved consistency in the apparatus, we have now attempted a list of values as in Table I. Here silica stands for the datum level so that other materials, all of them elements, give P.D. values, positive or

* Shaw and Leavy, 'Phil. Mag.', vol. 10, p. 809 (1930).

† Shaw, 'Proc. Roy. Soc., A', vol. 111, p. 341 (1926).

Table I.—Present results (see columns 1 and 2), compared with other lists, columns 3, 4, 5, 6, 7, the last four of which are reversed in sign. As regards column 3 the usual list has Zn positive, Pt negative, since the first metal gives up anions most readily, but throughout the present work we consider the charge *remaining* on the metal; so we put Zn positive to Pt in this column.

1.	2.	3.	4.	5.	6.	7.
Tribo-electric series.	Volts.	Inverted electro-chemical series.	Volta series.	Contact potential series.	Combination of Volta thermo-electric and thermionics.	Tribo-electric series, Shaw.
+		+	—	—	—	—
Au	+17	Pt	Pt	Pt	Pt	Pt
Pt	+15	Au	Au	Ag	Ag	Au
Ag	+14	Sb	Cu	Au	Cu	
Cd	+10	Ag	Ag	Cu	Bi	Ag
Cu	+ 9.3	Cu	Fe	Bi	Sn	Sn
Bi	+ 8.5	Bi	Bi	Pb	Fe	Co
Ni	+ 7.8	Sn	Sb			Sb
Pb	+ 7	Cd	Pb			Cu
Fe	+ 4.8	Pb	Cd			Ni
Cr	+ 3.3	Ni				
Co	+ 1.4	Co				Fe
Tl	+ 0.6	Fe				
Silica.						
Al	— 0.1		Sn	Sn	Zn	Zn
Sb	— 0.6		Zn	Zn	Al	Al
Zn	— 2	Zn	Al	Al		Cd
Sn	— 7.3	Al				
Se	— 7.7					
—		—	+	+	+	+

negative, against silica. We assume that the relation of the elements to silica indicates their relation to one another.

The values appear as volts attained at one stroke on the combined inductor, wiring, and attached electrometer. These results are specifically true only on this apparatus, used as described; since they depend on the capacity of this apparatus and on the present way of using it.

Table I, columns 1 and 2 show the noble metals at the head, the positive end, the base metals and metalloids at the bottom, the negative end; and the group Fe, Ni, Co, Cr, midway; the three heaviest elements Bi, Pb, Tl (see Table II, column 2) do not seem to fit well in the scheme. The list in Table I

Table II.—The list of elements in triboelectric order placed against their atomic numbers and outer shell electron numbers.

1. Triboelectric series.	2. Atomic number.	3. Number of outer shell electrons.
+		
Au	79	1
Pt	78	2
Ag	47	1
Cd	48	2
Cu	29	1
Bi	83	5
Ni	28	2
Pb	82	4
Fe	26	2
Cr	24	1
Co	27	2
Tl	81	3
Al	13	3
Sb	51	5
Zn	30	2
Sn	50	4
Se	34	6
—		

column 1, suggests comparison with other atomic relations. Consider these in turn :—

I.—The relation of the triboelectric series to volta contact series is of historical interest ; for, since the time of Helmholtz (1879), and following his lead, many physicists have held that the two effects are identical in origin ; so that knowing, say, the volta series, the tribo-series follows without further trouble. For reasons already given (Shaw and Jex, *loc. cit.* (1926)) we think this theory, while basically true, needs considerable amplification, since a complex dynamical and thermal process like friction must, under ordinary conditions, contain many factors which do not enter the simple statical volta effect. But the present method of rubbing, as explained above, has the merit of avoiding certain dynamical and thermal complications ; so we may hope to find some correspondence between our list and the well-known volta lists. Unfortunately no quantitative values for the volta series exist except for a few elements, as the modern method of exhaustive out-gassing in measuring contact potentials has only been applied to a few metals.

In Table I, column 4, is a volta list taken from Winkelmann's "Handbuch" ; in column 5, a contact potential list from the 'International Tables,' vol. 6 ; in column 6 a combination list, Shaw and Jex, *loc. cit.* (1926) as shown ; and in column 7 a list from one of our former papers. These four lists are grouped

together since they all have negative elements at the head, positive at the base.

Contrasting with columns 4, 5, 6, 7 there is in column 3 an electrochemical list showing the chemical relation of the elements. This list with the tribo-list, column 1, is alike in having positive at the head, negative at the base. Thus we have the extraordinary fact that the six lists are in rough agreement; for the three groups of noble metals, iron group and base metals are in the order stated, but four of the lists are one way up, the other two are reversed. Consider the two classes.

Class I.—Base metals positive (columns 4, 5, 6, 7).

Here the volta static results agree with the triboelectric results, when the metals are not rubbed on solids of the glass family and mineral silica compounds.

Class II.—Noble metals positive (columns 1, 3).

These two lists have this in common: the metals are brought into contact with inorganic acids, bases, and salts. For instance, zinc is negatively charged as against platinum when these are opposed in a voltaic cell in presence of an acid; and the same relationship holds between zinc and platinum when they are compared against silica and its compounds. Some physical chemists hold that glass is a solid electrolyte with sluggish anion and active kation (see * p. 259). An apparatus is described in which a glass globe, 0.1 mm. thick, containing salt solution and surrounded by another salt solution is used, and a P.D. found between the glass and the solutions. Bearing this in mind, we suggest that in class I the important factor is the simple surrender of electrons across the interface of metals, to that metal having the stronger positive field, there being no chemical action to complicate the issue. But in class II the decisive factor, as between the noble and base metals, seems to be Nernst's solution pressure† which is greater in base metals than in noble metals, and renders the former relatively negative. Glass having free silica or silicates seems to simulate, in feeble degree, the action of inorganic acids or salts, rendering the base metals (which possess strong solution pressure) negative to the nobler ones.

* Freundlich, "Colloid and Capillary Chemistry" (Methuen).

† The P.D. between two metals placed in the same electrolyte is:

$$V_2 - V_1 = R_1 T \log P_1 - R_2 T \log P_2,$$

where P_1, P_2 are solution pressures, p_1, p_2 osmotic pressures, and R_1 and R_2 values of the gas constant according to valency. To apply this effect with such a substance as glass involves the passage of the metal ions into the glass, a surface affair, and therefore soon completed.

In a former paper of this series* we showed that certain materials zinc, silk, filter paper, cotton, glass, can be arranged in a closed curve with potentials falling all the way. Here the glass acts like acid on the zinc and builds up potential as in a voltaic cell. This paradox is in agreement with the reversal of metals in Table I.

If, as we assume, the reversal of our list is attributable to the action of silica and its compounds, it might be thought better to avoid silica, and confine ourselves to metals and to organic non-conductors, and thus obtain truer comparison between the tribo-series and the volta series. But purity of surface is the first essential in this work, and it seems hopeless, as we have shown before (Shaw and Jex, *loc. cit.*, 1926) to try to obtain clean organic surfaces of constant action, even in the case of highly purified cotton and silk fabrics. On the other hand, members of the glass family, being refractory, can be well cleaned in a Bunsen flame; and being non-conductors they are suitable for rubbing against the metals. For these reasons, silica, the simplest member of the glass family is selected.

Leaving aside, for the time being, comparison with the volta series, let us turn to certain established data which may throw light on the genesis of our charges.

II. In Table II the series in column 1 is compared with atomic numbers in column 2. In general the heavy atoms are high in the list and the light ones are lower down (but Tl, Sb, Sn are low compared with other atoms of like weight).

III. Turning to column 3, which shows the number of electrons in the outermost, incomplete, quantum shell in Bohr's Table of the elements, we observe that small numbers tend to the top, and large ones to the bottom of the column; and this principle is specially noticeable for the shell numbers 1 and 2. (Bi and Pb, however, seem out of place on this plan.) This is in accordance with the well-known fact that elements like the alkalis with a small number of outer shell electrons are prone to share them with atoms like the halogens which are rich in electrons, *i.e.*, the atom with few outer shell electrons surrender them easily so as to establish a stable ring group.

IV. The ionisation potentials of the elements might be expected to be high for the metals low in our triboelectric list; but there seems to be no relation between the two effects. It must be remembered that ionisation potentials apply strictly to *free* atoms whereas in solid surfaces, such as we use,

* Shaw and Jex, 'Proc. Roy. Soc., A, vol. 118, p. 110 (1928).

the atoms are not free ; in the present work we have to deal with unknown molecular, as well as known atomic, properties ; and looking back to our remarks on atomic number and shell numbers, Table II, it may be that the lack of correspondence between the lists may be due in part to the fact that atomic properties are not the only ones to be considered.

V. The thermo-electric differences between metals are well known ; but they shed no light on our subject, for, as Coehn* says, "it is not yet possible to show any relation between the thermo-electric and other properties of a metal."

Theory.—The simple contact theory of frictional electricity is this. Let the area of surface rubbed have length l , breadth b . Of this area (since the surface is rough compared with molecular dimensions) only a fraction has molecules actually in contact. The true contact area is therefore $n \cdot b \cdot l$. The charge produced $Q = C \cdot v = (nbl/4\pi d)v$, where C = capacity of the opposed surfaces, v the potential set up by electrical exchanges due to contact and rubbing and d = distance between the surfaces.

When the surfaces are separated, the capacity falls to c and the P.D. rises to V , so $(nbl/4\pi d)v = cV$. To use this theory the breadth of contact b as well as its length l must be the same for all rods. But the rods vary in diameter and material so our experimental values all have to be corrected by Hertz's deformation equation for crossed cylinders.† In this way the column 1, Table I, was obtained by correction of the readings. Of these factors b , l , V are measurable, d = distance between atoms. In the case of a perfect insulator well away from conductors, c might be estimated from the known area bl ; but the term n is indeterminate, and it varies according to the hardness and roughness of the surfaces and according to their pressure on one another. Not knowing n we cannot find v . This is one technical difficulty in the subject. There are many others inherent in this surface work as we shall now see. Without attempting a mathematical treatment we may point out some of the factors which should be included in a full theory.

In a previous paper (Shaw and Jex, *loc. cit.*, 1928) a simple additive expression for P.D. obtained is :—

$$V = E - P \pm S - M \pm A - T, \quad \text{II}$$

where each term stands for potential, caused by the effect in question to be

* 'Handbuch der Physik,' vol. 13, p. 334.

† Shaw, 'Phil. Mag.,' vol. 1, p. 205 (1901).

added to the common stock. We will now modify and enlarge the expression to :

$$V = (E + P) + (S + T) + (M + A + O) + (n + r + p), \quad \text{III}$$

the new terms being in small type, and the addition being strictly algebraical. The terms are :—

- E = electron surrender,
- P = solution pressure (Nernst),
- S = surface strain before or during the stroke,
- T = temperature above normal during the stroke,
- M = moisture film,
- A = acid or alkali film,
- O = organic film,
- n = a fraction (defined above),
- r = streak, say metal, on silica, left by rubbing,
- p = electric separation, arising when the surfaces part.

Each term in equation III represents the joint action of the two opposed surfaces.

The terms in III are arranged in four bracketed groups. The first ($E + P$) stands for the electric separation occurring when the pure surfaces meet. This bracketed term therefore is the volta contact effect ; and in a great many cases P vanishes leaving only E. The last group ($n + r + p$) includes terms which cannot be eliminated by any cleansing of surfaces or by any device for rubbing them ; n and p are indeterminate variables ; r shows itself in some cases, as when metals leave a trail if rubbed on certain organic or inorganic solids ; but it is not apparent, and may not occur, in other cases such as the rubbing of cotton on glass. If a streak occurs it becomes involved in p and necessarily has a considerable effect on the ultimate charge ; for when the surfaces separate the rupture takes place not between, say, metal/silica, but between metal/metal-silica and the potential arising is electrokinetic not thermodynamic.* The second bracket term in III includes ($S + T$). These two troubles, S and T, may be reduced to small account by care in method of diagonal rubbing. The third group ($M + A + O$) includes films which one may hope to remove, or at least reduce to negligible influence, by avoiding moist air, chemical treatment of the surfaces and organic dirt, respectively.

* Freunlich, "Colloid and Capillary Chemistry," p. 252.

If we can remove the second and third bracket terms by the methods suggested, (III) reduces to

$$V = (E + P) + (n + r + p), \quad \text{IV}$$

so that we still encounter n , r , p three unknowable terms; and even if in some cases, r and p are counted out, n remains, an unknown addition to the volta effect ($E + P$). Thus we must expect to find considerable disparity between the volta list and the triboelectric list under the most favourable conditions.

There is another reading to equation III. Though it applies strictly to electric charges, the terms on the right side must influence the force required to slide one solid over the other. The tangential force in friction is made up of (1) the shearing force in separating two surfaces bound electrically, (2) gravity force, the over-riding of lumps, which is increasingly small as the surfaces become smoother. The point of this comment is that friction like triboelectricity has all the complication involved in equation III.

Vacuum Work.—Looking at equation III we see that certain films, especially water film, can be eliminated by work in vacuum, preferably with heat. Many attempted experiments *in vacuo* have been made.* But to conduct a series of rubbing experiments where considerable force is required, and to renew and anneal surfaces between the rubs, demands joints and stuffing boxes which would inevitably leak and emit vapours, so that under such conditions high vacuum and clean surfaces would be unattainable. Then we have the troubles incurred in trying to get good quantitative electrical readings from inside the vacuum tube. The technical difficulties appear to us well-nigh insuperable. If they could be overcome, and the film troubles avoided, the equation III would still contain several disturbing factors; so the labour of working in a vacuum in our opinion would be in vain.

Recent Research.

Notable experimental work of the last two decades is to be found in a series of papers.† Physicists, so active on other topics, have kept at respectful

* Shaw and Jex, 'Proc. Phys. Soc.,' vol. 39, p. 5 (1927); Knoblauch, 'Z. phys. Chem.,' vol. 39, p. 225 (1902); Coehn and Cuss, 'Z. Physik,' vol. 29, p. 186 (1924); French, 'Phys. Rev.,' vol. 9, p. 151 (1917).

† Shaw, 'Proc. Phys. Soc.,' vol. 27, p. 208 (1915); 'Proc. Roy. Soc.,' A, vol. 94, p. 16 (1918); Shaw and Jex, 'Proc. Roy. Soc.,' A, vol. 118, pp. 97, 108 (1928); *ibid.*, vol. 122, p. 49 (1929); Shaw and Leavey, 'Phil. Mag.,' vol. 10, p. 809 (1930); Richards, 'Phys. Rev.,' vol. 16, p. 290 (1920); Vieweg, 'J. Phys. Chem.,' vol. 30, p. 865 (1926); Coehn and Lotz, 'Z. Physik,' vol. 5, p. 242 (1921); French *loc. cit.*; McClelland and Power, 'Proc. Roy. Irish Acad.,' A, vol. 34, p. 40 (1918); Macky, 'Proc. Roy. Soc.,' A, vol. 119, p. 107 (1928).

distance from this subject. Perhaps they are conscious that this, the oldest and easiest way to produce electricity, proves on investigation to be the most involved. But in any case it can hardly be that such a common, and therefore practically important, effect as friction, will much longer suffer from neglect.

A recent contribution to theory is made by P. W. Burbidge, in "A tentative Theory of Frictional Electricity."* The theory is based on the principle that all solid and liquid surfaces possess reacting fields of force. Two cases are considered :—

(1) A crystal of the cubical system has every particle carrying the same charge at equal distance from its neighbours. The charges and spaces are known, so that the field at any distance of atomic dimensions can be calculated. When two surfaces are brought into contact, if their materials are different their fields also are different. And the forces on electrons on the opposed surfaces being different, there will be a greater stream of electrons in one direction than in the opposite. The supposition made is that the fields are strong enough to detach electrons from the opposed surface across the interface.

(2) A dielectric consists of polar molecules (Debye). These dipoles, in the powerful fields which exist on all free surfaces become oriented and then, from these oriented dipoles a transfer of electrons ensues, as in the case of a crystal.

Now Millikan has succeeded in drawing electrons from cold metals by the use of fields of 10^6 volts/cm. By using known data for cases (1) and (2) Burbidge calculates that their surface fields are of the same order (10^6 volts/cm.). Thus this theory provides us with a quantitative basis for the production of charges by contact, at least as regards simple surrender of electrons.

Summary.

Experiment reveals the complexity of this subject. There are ten factors, at least, concerned in the process of rubbing two solids. These factors operate both in the electrical process of producing charge, and in the dynamical process of friction, between the solids. If the potential due to each factor be taken as a distinct term, we can write for the net potential due to rubbing :

$$V = (E + P) + (S + T) + (M + A + O) + (n + r + p).$$

The first bracketed group stands for the electric separation when pure surfaces meet, *i.e.*, for the volta contact effect. The second group, which involves

* 'Trans. Proc. N.Z. Inst.,' vol. 59, p. 663 (1928).

strain and temperature, depends on the method of rubbing. The third group includes effects of surface contamination. The last group includes factors inherent in the process of rubbing. Unlike groups two and three, these factors operate, however carefully or cunningly the rubbing be conducted. Thus it comes about that the charges produced by *friction* are not simply those set up by mere *contact*, as in the volta effect.

This paper is experimental. Working on pure elements and the simple refractory material, silica, with improved apparatus and methods, the relation is examined between the volta effect and the triboelectric effect. The two effects prove to be somewhat similar, but there are disparities. The conclusion reached is that, of the four groups in the equation the first is, at least in the case of these specially chosen and carefully prepared materials, paramount for most elements, but not for all.

An X-Ray Investigation of Normal Paraffins near their Melting Points.

By ALEX MÜLLER.

(Communicated by Sir William Bragg, F.R.S.—Received July 12, 1932.)

[PLATE 13]

Introduction.

The present paper is a continuation of previous work* on the thermal expansion of normal paraffins. It is confined to an investigation of the lattice changes which take place in the range between room temperature and the individual melting points. The observations are rather complex and certain difficulties in the interpretation of the X-ray photographs arise which make further investigations necessary. Such work is now in progress. There were, however, several unexplained points in the last paper which are now clear and dealt with in the present work.

Range of Materials.

The investigation was carried out with 15 specimens ranging from $C_{18}H_{38}$ to $C_{44}H_{90}$. The method of their preparation has been described previously.

* 'Proc. Roy. Soc.,' A, vol. 127, p. 417 (1930).

Experimental.

The problem of the present research was to measure the lattice dimensions at any given temperature between 20° C. (room temperature) and the melting point of the individual substance, the highest being about 90° C. The recording of the lattice dimensions was obtained with the aid of X-ray photographs. It was essential to keep the temperature of the substance constant to a fraction of a degree during the exposure to the X-rays, and it was also desirable occasionally to vary the temperature in small steps of a few tenths of a degree.

The investigation of one single specimen involved sometimes 10 to 20 temperature steps. It was therefore desirable to reduce the time for the temperature adjustment as much as possible. The small thermostat in which the substance had to be kept during the X-ray exposure was heated to the required temperature by a constant flow of preheated water. The water was heated in a small copper tube boiler with a gas burner. There were two ways of adjusting the temperature; one consisted in regulating the gas flame, the other in varying the flow of water. There were three thermometers in the circuit. The first was between the boiler and the thermostat, the second in the thermostat and the third at the outlet. With this arrangement it was possible to estimate the maximum temperature drop in the thermostat itself. This drop did not exceed 1° at the highest temperature recorded and the actual readings were taken at the thermostat thermometer. There were three sets of these used with overlapping scales and each reading to 1/10°.

The thermostat itself was made from a solid copper rod. It weighed about 1 kilogram. It had hollow spaces through which the water circulated and slots through which the incident and the reflected X-ray beams entered and emerged. These slots were either covered with thin nickel foil or cellophane, which stopped any air circulation. The specimen itself, spread on a flat surface at the end of a copper rod, was in the centre of the thermostat. This rod was tapered and fitted tightly in a corresponding seating in the copper block. The thermometer bulb was completely surrounded by copper and close to the specimen. The whole thermostat was supported by three adjustable screws on the table of an X-ray spectrometer. The flat surface on which the substance was spread went through the axis of rotation of the spectrometer. The layer of substance covered about 1/2 sq. cm. and was about one to two-tenths of a millimetre thick. It was obtained by melting a small quantity of the substance on the flat portion of the copper rod.

The spectrometer table carrying the thermostat was oscillated through

about 5° during the exposure. The flat top of the thermostat had a circular scale and the copper rod had a pointer. This made it possible to set the surface of the layer at any required angle relative to the X-ray beam, and to obtain a certain degree of focussing of the reflected rays for a certain region of the spectrum.

The reflected rays after emerging from the thermostat had to pass through a long narrow slot in a shield behind which was a film drum. This drum was turned through a small angle after each exposure, and in this way a whole series of photographs were recorded on the same film. It was easy to measure the shift of any of the recorded lines relative to a reference line, and this was all that was required for the measurement of the expansion. The geometrically correct focussing was, of course, not obtainable for more than one line, but since the actual reflecting range covered only about 10° the lack of correct focussing was hardly appreciable.

The distance between the focus of the X-ray generator and the specimen was about 8 cm. and usually 5.95 cm. between the axis of the spectrometer and the film. Calibration photographs were sometimes taken at twice this distance. A considerable time saving was obtained by the use of an X-ray generator with a rotating anode. The generator had a copper anode and ran as a rule at about 30 KV. and 120–150 milliamps. The average time of exposure was only a few minutes. The temperature adjustment took not more than 5 minutes. The X-rays were, as a rule, filtered by nickel foil and the wave-length used was 1.539 Å.

Results.

There are three types of paraffin crystals known so far :—

- (1) The normal form. The chains are packed in a prismatic cell of rectangular cross section. The chains are perpendicular to the base of the cell.
- (2) A form of lower symmetry. The cross section is not rectangular and the chains may be tilted relative to the base.
- (3) A form which has a rectangular cross section (or at least very nearly rectangular cross section) and chains which are tilted relative to the base. This form has been found by Piper and Malkin.

It is found that $C_{18}H_{38}$, $C_{20}H_{42}$ and $C_{22}H_{46}$, when observed at room temperature, belong to the least symmetrical form. $C_{19}H_{40}$ and $C_{21}H_{44}$ show the normal structure and from $C_{23}H_{48}$ up to $C_{34}H_{70}$ both odd and even numbered substances exist in the normal form provided they have previously been molten.

The third form is, as a rule, only observable when the crystals are obtained directly from a solvent. Exceptions to these rules are observed most frequently with substances above $C_{30}H_{62}$. The exact conditions for the occurrence and the stability range of these crystal forms are a subject for further investigations; their study is outside the scope of this paper.

Previous work on the expansion of normal paraffins has shown that the length of the chain axis depends much less upon the temperature than the length of the two other axes. These observations have been confirmed in the present work. The expansion of the c axis (chain axis) is too small to be measured with the apparatus used in the present work. The investigation is essentially confined to the expansion of the two axes which are in a plane perpendicular to the chain axes, and the discussion of the numerical data will mainly deal with those obtained from the normal form.

The planes which are used in this paper for the determination of the " a " and " b " axis, *i.e.*, the two axes which are perpendicular to the chain axis, are all in a zone which contains this chain axis. The actual planes are: 110, 200, 210, 020 and 310, the first two being by far the strongest. These two spacings are indicated in the diagram fig. 1.

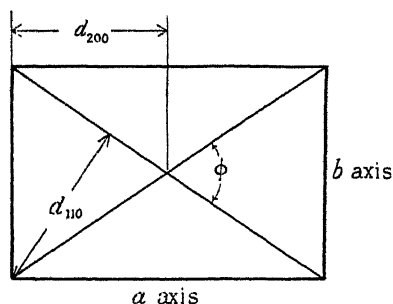


FIG. 1.—Cross section of unit cell (normal form).

The equation connecting the spacings d_{hko} with the axes and indices is

$$\frac{1}{d_{hko}^2} = \frac{h^2}{a^2} + \frac{k^2}{b^2}. \quad (1)$$

There are essentially two unknown quantities which have to be calculated from the observations, namely, " a " and " b "; the indices being known from previous work. Observations were, as a rule, made on the five spacings mentioned above. The method of the least squares was applied in the following form:

$$\Sigma \left(\frac{1}{d_{hko(\text{obs.})}^2} - \frac{h^2}{a^2} - \frac{k^2}{b^2} \right)^2 = \text{minimum}. \quad (2)$$

The choice of this expression is arbitrary to a certain extent. It has the advantage of making the calculations simple and to give comparatively higher weight to the more accurate observations at large deflecting angles.

The following table gives the result of measurements on a number of paraffins at room temperature. The axes are calculated with the aid of equation (1) and (2) from direct measurements of the five spacings.

Table I.

Number of carbon atoms.	Temperature.	a .	b .	$ab/2$.	ϕ ($\tan \phi/2 = b/a$).
	° C.	A.	A.	sq. cm.	° ' "
19	19.5	7.55	5.01	18.9×10^{-16}	67 10
23	18.4	7.43 _s	4.97	18.2×10^{-16}	67 28
24	18.7	7.41	4.94	18.3×10^{-16}	67 20
25	18.7	7.41	4.96	18.4×10^{-16}	67 32
26	19.1	7.41 _s	4.94	18.3×10^{-16}	67 22
27	19.6	7.40	4.93	18.2×10^{-16}	67 22
29	19.5	7.42	4.94	18.3×10^{-16}	67 16
30	19.5	7.33	4.92	18.2×10^{-16}	67 44
31	19.6	7.40	4.93	18.2×10^{-16}	67 15
34	19.3	7.40	4.95	18.3×10^{-16}	67 30
44	20.2	7.33	4.93	18.1×10^{-16}	67 52

(These measurements were taken with specimen enclosed in thin-walled glass tubes of about $\frac{1}{2}$ millimetre diameter. The glass tubes were centred in the axis of the thermostat.)

$a, b/2$ is the cross section area occupied by one molecule. Since the chain axis is treated as independent of the temperature in this work, the cross section is proportional to the molecular volume at any given temperature.

The following table shows typical expansion data from a specimen of $C_{23}H_{48}$. The data in this table are calculated from the five observed spacings at each of the given temperatures. The cross section of the specimen at room temperature is taken as a standard. The actual photographs are reproduced on Plate 13. The 110 and the 200 reflection draw closer and closer together and become indistinguishable from each other as the temperature approaches the melting point. The angle ϕ between the diagonals, fig. 1, becomes 60° , showing that the structure changes into hexagonal close packing.

The numerical data of Table II are shown in the following graphs.

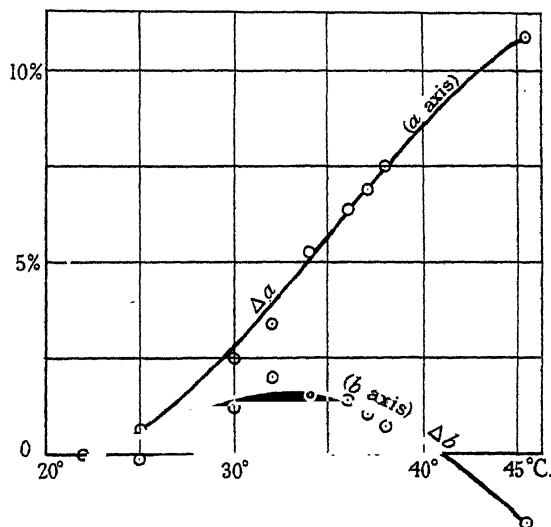
The transition from the less symmetrical form into the hexagonal close packing is a continuous function of the temperature.

Table II. $n - C_{23}H_{48}$.

Temperature.	Δa .	Δb .	ϕ .	
°			° /	
22	± 0	± 0	66 52	
25	-0.1	0.63	67 10	
30	+2.5	1.2	66 0	a_t observed length of a
32	3.4	2.0	65 58	axis at temperature t .
34	5.3	1.5 ₆	64 48	
36	6.4	1.4	64 10	b_t observed length of b
37	6.9	1.0 ₆	63 48	axis at temperature t .
38	7.5	0.7	63 20	
45.3	10.9	-1.8	60 0	

$$\Delta a = \frac{a_t - a_{22}}{a_{22}} \cdot 100$$

$$\Delta b = \frac{b_t - b_{22}}{b_{22}} \cdot 100.$$

FIG. 2.—Percentage expansion (obs.) of a and b axis (Table II).

$C_{21}H_{44}$ behaves in the same way as $C_{23}H_{48}$. $C_{19}H_{40}$ differs from the two in so far that the substance melts before the symmetrical state is reached, Plate 13. The even numbers C_{18} , C_{20} and C_{22} at room temperature have a different structure compared with the odd numbers. C_{18} and C_{20} do not reach the state of hexagonal packing at the melting point, whereas C_{22} does.

A new phenomenon is observed with $C_{24}H_{50}$. The " a " and " b " axes alter their length as the temperature increases, but this change seems to be much smaller at first. When a certain temperature is reached there is a sudden change in the structure. This is easily seen on the photographs, Plate 13.

The appearance of the photographs alters. Out of the five lines there are only two left, the stronger of the two being only very slightly displaced relative to the 110 line before the transition. The second and weaker of the two shifts into a position about half way between the original 110 and 200 reflection. Occasionally it is found that the original 200 line still persists after the transition, but it is always much fainter than before the transition. Sometimes a number of faint lines are observed. These lines appear to be stronger if a glass tube specimen is used instead of a layer. On one of these photographs as many as six lines are present. All these lines are close to the original 110 and 200 reflection, their spacing being between 3.5 and 4.5 Å. The lines with larger reflecting angles seem to vanish. It is only after long exposures—over an hour—and at increased distance, that faint traces of large-angle reflections appear. The existing data are not complete enough to give a picture of the structure change which takes place at the transformation point.

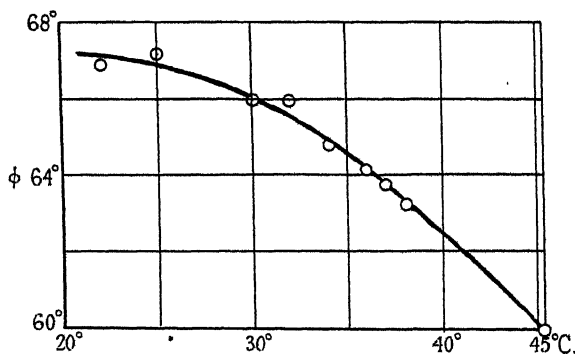


FIG. 3.—Angle ϕ (obs.) between diagonals of cross section (Table II).

The photographs become simpler as the temperature increases. The faint lines vanish very soon and the two remaining stronger lines draw together and behave in the same way as those observed with C_{21} and C_{23} . Near the melting point there is only one line left in the same position as the corresponding lines of C_{21} and C_{23} .

This fact seems to suggest that whatever the structure of the material may be near the transformation point, the final arrangement is again the hexagonal packing. The following table gives the observed spacings corresponding to that single line, which incidentally has already been observed in previous work (*loc. cit.*). The two last columns contain the cross section and also the distance of nearest approach measured between chain axes. The angle ϕ is now 60° for all the specimens in the table.

Table III.—Data on Hydrocarbons in the Solid State near their Melting Point.

Number of carbon atoms.	Melting points.	Observed spacings.	Cross section per molecule.	Distance of nearest approach.
	°	Å.	sq. cm.	Å.
22	44	4.11	19.6×10^{-16}	4.76
23	46	4.17	20.1×10^{-16}	4.81
24	51	4.12	19.6×10^{-16}	4.76
26	58	4.10	19.4×10^{-16}	4.73
27	61	4.06	19.0×10^{-16}	4.69
29	64	$\left\{ \begin{array}{l} 4.10 \\ 4.03 \end{array} \right\}^*$	19.2×10^{-16}	4.68

* Nearly hexagonal.

Photographs of the substances in the molten state show a diffuse band instead of the well-defined lines characteristic of the solid. A few of the observed spacings of molten substances are tabulated below.

Table IV.—Data on Hydrocarbons in Liquid State at their Melting Point.

Number of carbon atoms.	Observed spacing.
	Å.
18	4.6
19	4.6
20	4.5
24	4.6
30	4.6

The range of substances which actually go into the hexagonal form seems to be limited. This state of highest symmetry is observed for the numbers 21, 22, 23, 24, 25, 26, 27. 29 begins to show a slight deviation. In C_{30} the two lines are clearly separated at the melting point, and the same holds with the still higher members so far as they have been observed. The same phenomenon is observed with C_{20} and the lower members of the series.

The abrupt transition which occurs with C_{24} is also found for all the higher members. The transition temperature gets closer and closer to the melting point and is almost indistinguishable from it with $C_{44}H_{90}$.

The following table contains the transition temperatures. They were obtained by examining the X-ray photographs taken at different temperatures. No attempt has been made to determine all the transition points very accurately. In the case of C_{26} the temperature steps were narrowed down to fractions of a degree and corresponding X-ray photographs were taken. It was found that the change of the X-ray photograph took place within a few tenths of a degree.

Table V.—Transition temperatures.

Number of carbon atoms.	Transition temperatures between	Melting points.	Difference.
	°	°	°
24	40–41	51.2	10
26	45.5–46	58.0	12
27	48.5–49.3	61	12
29	56.7–57.1	64.4	7
30	58.0–58.3	66.6	9
31	60–64	68.4	6
34	67–68.9	72.8	5
44	85.6–86	86.4	$\frac{1}{2}$

These transition temperatures have been measured by Garner* and more recently by Piper† and his collaborators. The present data, for which less accuracy is claimed, agree within the limits of experimental error.

It has been mentioned that the exact interpretation of the X-ray photographs obtained with specimens at the transformation point is not feasible at this stage of development. The large number of lines suggests that the material consists of a mixture of several crystal forms, most of which, however, seem to be stable over only a small range of temperatures. The fact that two of the lines are predominant and behave at higher temperatures like those of C_{21} and C_{23} leaves room for the following speculation.

Assuming that the bulk of the material has essentially the same arrangement as that before the transition has taken place, we may tentatively give the two lines the indices 110 and 200. The new structure differs from the old, in that the axes have different lengths. The data calculated from observations made with C_{29} and C_{24} are :

Table VI.— $C_{29}H_{60}$ Melting Point 63°. Expansion Measurements.

Temperature.	Δa .	Δb .	ϕ .
°			°
19.0	± 0	± 0	67 15
30	0.2	0.4	67 20
40	0.6	0.6 ₅	67 20
51	1.2	0.7	67 0
56.7 } trans-	1.6	0.7 ₅	66 50
57.1 } formation	6.7	± 0	64 20
63	10.5	-0.2	61 20

$$\Delta a = \frac{a_t - a_{19}}{a_{19}} \cdot 100 \quad \Delta b = \frac{b_t - b_{19}}{b_{19}} \cdot 100.$$

* 'J. Chem. Soc.,' p. 1533 (1931).

† 'Biochem. J.,' vol. 25, pp. 2072–2074 (1931).

The table for C_{29} is arranged identically with that of C_{23} . Up to the transition point the above hypothesis is not involved. The data of this table are plotted in the following graph.

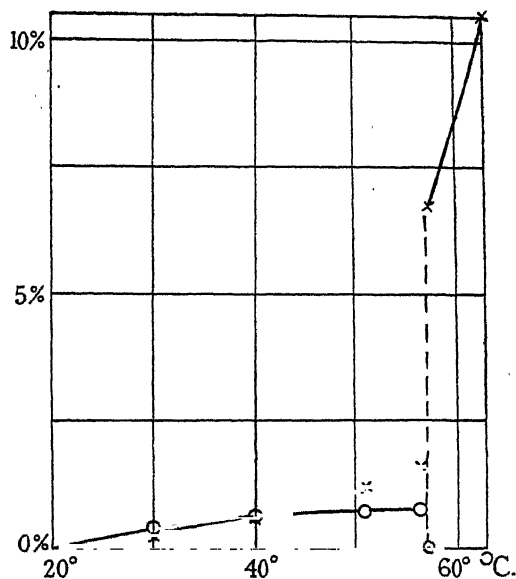


FIG. 4.—Percentage expansion (obs.) of a and b axis (Table VI):

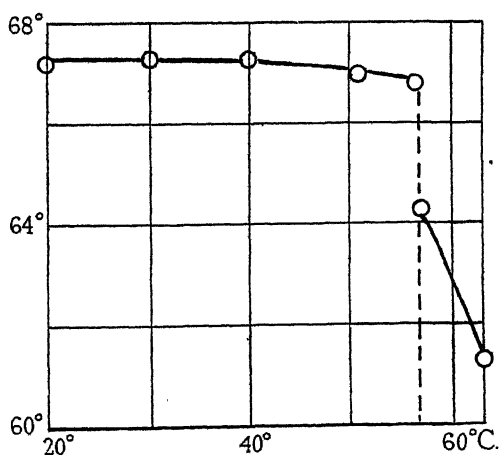


FIG. 5.—Angle ϕ (obs.) between diagonals of cross section (Table VI).

The following table gives the observed spacings of the two strong lines which remain after the substance has gone through the transition.

Table VII.— $C_{24}H_{50}$ Melting point: 51.2° C. Observed Spacings of the two Strong Lines.

Temperature.	d strong.	d weak.
°	A.	A.
40.0	4.12	3.71
41.4 ₅	4.16	3.91
41.8 ₇	4.14	3.93
42.5	4.12	3.94
42.9 ₅	4.12	3.97
43.4 ₅	4.11	4.01
43.9 ₂	4.09	4.08
45.4	4.10	
47.9		
49.3		
	4.11	
	4.12 ₅	

Using the above-mentioned hypothesis, which assumes that the two lines are the 110 and 200 of the normal structure, the following data are calculated from Table VII.

Table VIII.—Lattice Dimensions of the Hypothetical Structure of $C_{24}H_{50}$ above the Transition Temperature.

Temperature.	a .	b .	ϕ .	$ab/2$.
°	A.	A.	°	sq. cm.
Transition { 40.0	7.43	4.96	67 26	18.4×10^{-16}
41.4 ₅	7.82	4.92	64 16	19.2×10^{-16}
41.8 ₇	7.86	4.87	63 34	$19.1_5 \times 10^{-16}$
42.5	7.88	4.83	63 0	$19.0_5 \times 10^{-16}$
42.9 ₅	7.95	4.82	62 30	$19.1_5 \times 10^{-16}$
43.4 ₅	8.03	4.79	61 40	$19.2_5 \times 10^{-16}$
43.9 ₂	8.16	4.72	60 0	$19.2_5 \times 10^{-16}$
45.4	8.20	4.73	60 0	19.4×10^{-16}
47.9	8.22	4.74	60 0	19.5×10^{-16}
49.3	8.25	4.76	60 0	19.6×10^{-16}

It is interesting to note the slight contraction of the “ b ” axis, which is also found in Table II for C_{23} .

Discussion of Results.

The data of the last chapter show that the normal paraffins tend to become hexagonal at the melting points. In the range of C_{21} to C_{29} this state of high symmetry is actually reached. Substances outside this range melt before becoming hexagonal.

X-ray investigation has shown that the CH_2 groups in the paraffin molecule are arranged in a zig-zag chain, and that the chains have only two planes of symmetry intersecting in the chain axis. This is found from measurements made at room temperature. The present work shows that the molecules behave as if they were more symmetrical at the melting point.

The inert character of the paraffin molecule, its heat resisting properties and the fact that the chain length does not depend appreciably upon the temperature, suggests a considerable rigidity of the molecular structure. It seems therefore unlikely that the temperature should produce a radical change of the configuration of the carbon skeleton of the molecule.

The carbon chain is surrounded by hydrogen atoms. This hydrogen shell is more likely to be affected by temperature. It is almost certain that the temperature motion of the hydrogens tends to make the molecule more symmetrical at higher temperatures, but it is impossible to tell at the present stage whether the temperature agitation of the hydrogen molecules accounts for the observed increase of symmetry, or whether the zig-zag structure of the carbon chain is still predominant.

The molecule as a whole must perform oscillations under the influence of the temperature. These oscillations also tend to make the crystal more symmetrical. The moment of inertia of the chain molecule is smallest relative to the chain axis. The amplitudes of the oscillations round this axis may become very large at higher temperatures and the molecules may even perform complete rotations. They would then, on the average, have the symmetry of a circular cylinder and the hexagonal close packing would follow quite naturally.

It must, however, be borne in mind that there are limits to the rigidity of the molecule. The frictional forces increase with the chain length and there must be a state when torsional vibration sets in. How this distortion affects the structure, and at which chain length it becomes appreciable, is again impossible to predict from the existing data.

The crude model of a rotating molecule gives at least a qualitative explanation of the phenomena observed with C_{21} and C_{23} . It fails, however, to explain the transitions which are observed with the higher members of the series.

These sudden transitions may appear surprising at first sight. Considering the similarity of the structure, the continuity of the melting points and other physical properties, it does not seem obvious why with two consecutive members of the series, like C_{23} and C_{24} , one should have a transition point and the other not. A qualitative explanation of a similar case has already been given in a previous paper (*loc. cit.*). The argument is this.

The forces which keep the molecules in these crystals together can be divided into two parts. One part of the cohesion can be ascribed to an interaction between the chains, the other between the molecules which form the end groups. These two forces are in equilibrium in the crystal, and the first of the two must depend upon the length of the molecule. The sudden appearance of a transition indicates that the balance of the two forces is sensitive to a small alteration of the chain length at this particular point.

Some of the statements in the last section may be put in a more definite form. Let us consider a chain molecule "A" surrounded by its neighbours. The centres of the circles in the diagram are the intersections of the two rows of carbon atoms in the molecule. The chain axes are in the middle between

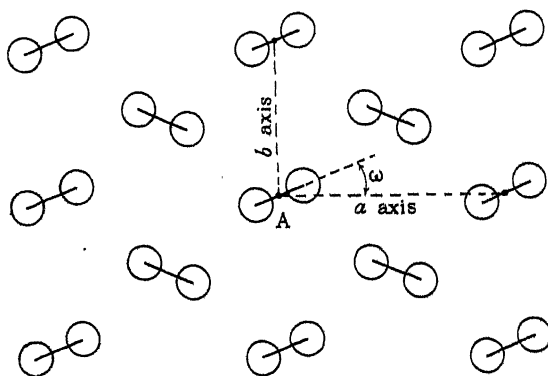
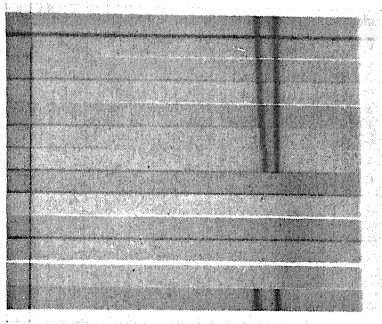


FIG. 6.

the two rows and are perpendicular to the plane of the paper. Let us further assume that the molecule is a rigid body of infinite length. The position of this body is then fixed relative to the crystal lattice by the angle ω . The average position of equilibrium may be called ω_0 .

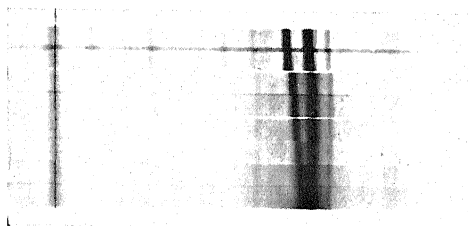
We now introduce the end groups. They are assumed to be so far apart from each other that their interaction is confined to pairs of adjacent layers. It is further assumed that these end groups produce only a small change in the original lattice, and that they are strongly linked to the chains. Supposing the chain forces were eliminated, the end groups would then arrange themselves in their own position of equilibrium—say $\bar{\omega}_0$.

A combination of the two forces is therefore likely to introduce two positions of equilibrium of the chain. The potential energy of the molecule plotted against the angle ω would have a shape as indicated in the diagram. According to the temperature, the molecules would be found more frequently in one or



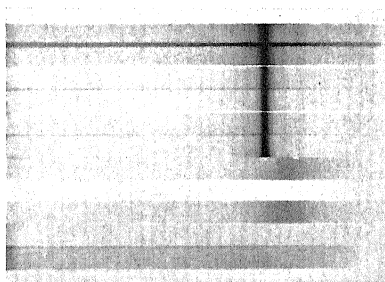
25.1—25.2
26.2—26.2
27.0—27.1
28.0—28.0
29.0—28.9
30.1—30.1
31.1—31.2
32.1—32.1
32.5—32.7
33.0—33.1
33.6—33.7
33.4—34.1
20.2—19.2

$n\text{-C}_{19}\text{H}_{40}$.

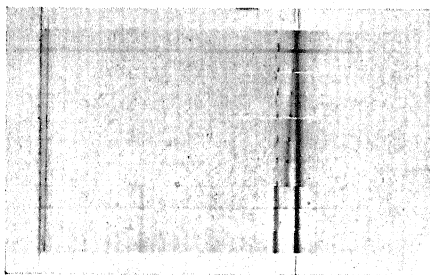


22.05—22.05
25.05—25.05
30.50—29.70
31.95—32.60
34.05—34.60
36.50—35.85
36.90—37.00
37.90—37.95

$n\text{-C}_{23}\text{H}_{48}$.

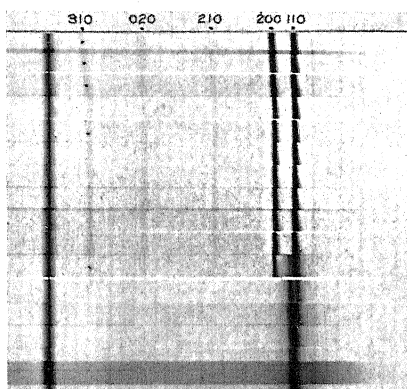


39.7 —39.7
40.0 —40.9
41.9 —41.95
43.0 —43.0
43.95—43.95
45.00—45.00
46.3 —46.2
46.95—46.90
47.6 —47.65
19.5 —19.2



43.90—43.95
43.50—43.40
43.05—42.85
42.55—42.50
41.85—41.80
41.50—41.40
41.00—40.95
40.05—40.00
38.90—38.75
34.85—

$n\text{-C}_{24}\text{H}_{50}$.



18.45— .45
30.00— .10— .10
39.95— .80— .80
48.80— .80—59.00
51.00—50.75—51.05
51.90—51.95—52.10
53.10—53.00
54.10—54.15
55.05— .35— .05
56.10—55.80—56.05
57.05—56.70—57.00
57.90— .90
59.90—60.10—59.75
60.95—61.15—60.85
61.95—61.70—61.90
62.85—63.00—63.00

$n\text{-C}_{29}\text{H}_{60}$.

the other potential valley. This representation, which is not meant to be more than a very crude picture, seems to be the most rational way to account for the observations made in the present work.

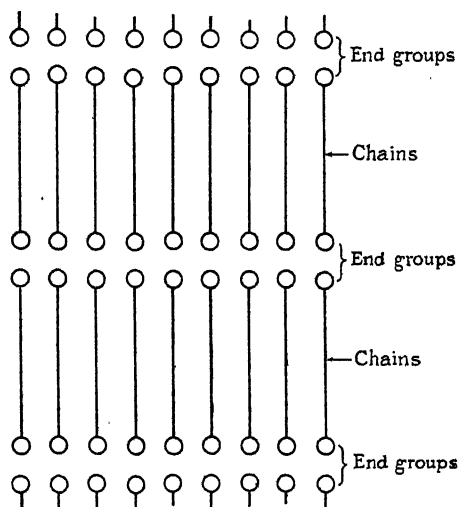


FIG. 7.

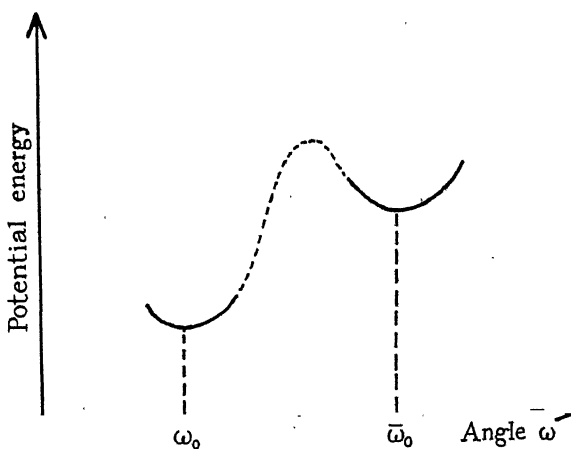


FIG. 8.

Further work is in progress which deals with a more detailed investigation of the disturbing effect of the end-groups, and the effect which is produced when atoms of different kinds are introduced into the chain lattice. It has already been found that oxygen atoms produce a contraction of the carrier lattice. These questions will be dealt with in a separate paper.

In connection with the last section it is of interest to use the data which Garner (*loc. cit.*) has published in his most interesting paper on the heat of crystallisation—and transition of several hydrocarbons. The following table gives the heat of crystallisation and transition of two hydrocarbons for which the changes in the lattice dimensions are known from the present investigation.

Table IX.

	$C_{24}H_{50}$	$C_{29}H_{70}$
Change of cross section area at transition point (sq. cm.)	0.8×10^{-16}	0.7×10^{-16}
Change of cross section area at melting point (sq. cm.)	2.1×10^{-16}	2.8×10^{-16}
Heat of transition (kg. cal./gr. mol.)	7.5	9.8
Heat of crystallisation (kg. cal./gr. mol.)	12.8	15.7

The figures for the energy changes are not those directly measured by Garner but they are obtained by interpolation from his measurements. The change in cross section at the transition points are taken from Tables VI and I and Table VIII of the present work. The change of the cross section at the melting points will be discussed in the next paragraph.

The point which these data are meant to illustrate is more clearly seen if the data of the last table are reduced to the chain element, *i.e.*, the CH_2 group. This is done by dividing the figures for the energy changes by the number of carbon atoms of the molecules. The average of C_{24} and C_{29} are used in the next table.

Table X.

$E_{Trans.}$	$\Delta_{Trans.}$	$\frac{E_{Trans.}}{\Delta_{Trans.}}$	Distance of nearest approach varies at transition point:
0.34	0.75	4.3	From 4.5 to 4.6 Å.
E_{Fusion}	Δ_{Fusion}	$\frac{E_{Fusion}}{\Delta_{Fusion}}$	At the melting point:
0.54	2.2	2.2	From 4.7 to 5.0 Å.

$E_{Trans.}$ = heat of transition
 E_{Fusion} = „ crystallisation
 $\Delta_{Trans.}$ = change in cross section p. mol. at transition point
 Δ_{Fusion} = „ „ „ „ melting point

} in kg. cal./gr. mol. CH_2 .
 } in 10^{-16} sq. cm.

This table gives an indication of the energy-changes involved at the transition—and melting point. The two ratios show the rapid decrease of the internal energy with increasing distance of the molecules (2.2 against 4.3).

The cross sections of the molecules in the liquid state which are used in the previous section are calculated from density measurements. These densities are taken from the International Critical Tables and show that the density

n = number of carbon atoms.	Density at melting point.
15	0.7761
16	0.7767
17	0.7766
18	0.7770
20	0.7778

at the melting point alters very little with the number of carbon atoms. The cross sections are calculated with the aid of the following expression :

$$\text{Cross section} = \frac{1.66 \times 10^{-24} (12.00 \times n + 1.008 [2n + 2])}{(1.253 \times n + 2.30) \times 10^{-8} \times 0.778}.$$

The numerator is the weight of the molecule and the denominator is the length of the molecule multiplied by the density (0.778). The length of the molecule is derived from earlier measurements on the long spacings of these hydrocarbons.*

The calculation gives the following result : The cross section area of a molecule in the liquid state at the melting point is 22.0×10^{-16} sq. cm. (average C_{20} to C_{30}). The distance of nearest approach between two chain axes, calculated under the assumption that the chains are hexagonally close packed in the liquid, is

$$\sqrt{\frac{2 \times \text{cross section}}{\sqrt{3}}} \sim 5.0 \text{ A.},$$

and the largest spacing in this hexagonal structure :

$$\sqrt{\frac{\text{cross section} \times \sqrt{3}}{2}} \sim 4.3 \text{ A.},$$

The observed spacings of the liquid at the melting point are all about 4.6 A. according to Table IV.

* These measurements were made on substances in the solid state. Owing to the negligible expansion of the chain axes and the comparatively small gaps at their ends, the use of these data is not introducing an appreciable error.

We therefore find :

Distance of nearest approach of hexagonally packed chains (calculated from density of liquid)	Observed spacing liquid at melting point	Spacing calculated from density (hexagonally packed chains)
5.0 A.	4.6 A.	4.3 A.

The distance of nearest approach is measured from chain axis to chain axis. The observed X-ray spacing lies about half way between the distance of nearest approach and the spacing calculated from the density of the liquid.

The writer wishes to express his appreciation to Sir William Bragg, O.M., F.R.S., and the Managers of the Royal Institution for their kind interest in the work.

Summary.

(1) The lattice dimensions of a number of hydrocarbons are measured at different temperatures between 20° and the individual melting points.

(2) It is found that $C_{21}H_{44}$, $C_{22}H_{46}$, $C_{23}H_{48}$, $C_{24}H_{50}$, $C_{25}H_{52}$, $C_{26}H_{54}$, $C_{27}H_{56}$ and $C_{29}H_{60}$ change from a state of lower symmetry into hexagonal symmetry. This state is reached when the substances are solid and near the melting point.

(3) The lower members of the series, *i.e.*, C_{18} , C_{19} and C_{20} , and also those above C_{20} , tend to approach hexagonal symmetry, but melt before they reach this state.

(4) C_{21} and C_{23} show a continuous change of the “*a*” and “*b*” axes with increasing temperature up to the melting point. C_{24} , C_{25} , C_{26} , C_{27} , C_{29} , C_{30} , C_{31} , C_{34} and C_{44} show abrupt transitions between room temperature and the individual melting points.

(5) Using Garner's data on the heat of crystallisation and transition, and the lattice changes measured in this work, it is found that the lattice energy decreases very rapidly with increasing molecular distance.

(6) Attention is drawn to the fact that paraffins of moderate chain length may be regarded as rigid rotators. Attempts are made to explain the transitions from general principles.

*Investigations in the Infra-red Region of the Spectrum. Part VII.—
An Infra-red Grating Spectrometer as a Double Monochromator.*

By A. B. D. CASSIE, and C. R. BAILEY, the Sir William Ramsay Laboratories
of Inorganic and Physical Chemistry, University College, London.

(Communicated by F. G. Donnan, F.R.S.—Received July 13, 1932.)

Part IV of the present series* describes a monochromator method for use in the infra-red region of the spectrum. The monochromator consisted of a single prism spectrometer with the absorption tubes placed between the telescope slit and the thermopiles, instead of the more usual disposition with the tubes between the Nernst filament and the collimator slit. The advantages of the method over the ordinary lay out are :—(i) double absorption tubes can be used and left permanently in position ; (ii) the introduction of stops can considerably reduce the quantity of scattered radiation of shorter wave-length reaching the thermopiles ; and (iii) only a small range of infra-red radiation traverses the gas under examination. The apparent disadvantage of the method is an increase in the number of reflections suffered by the infra-red beam, and a resultant loss in intensity ; in practice, however, it is found that this loss is small, and at shorter wave-lengths the same slit width was used as in the older method. As a consequence we decided to design a grating spectrometer for use as a double monochromator, and the present paper describes the instrument, with the method of use.

The Spectrometer.

The final arrangement is shown in fig. 1. The essential departure from the grating spectrometer constructed by Sleator† lies in the use of two mirrors in the spectrometer proper, instead of one. As was pointed out by Czerny,‡ in the simpler case, the error introduced in the image of the collimator slit on collimation of the beam will be approximately doubled when the reflected parallel beam is converged by the collimating mirror to form an image near the collimator slit ; if, however, two mirrors are used with the collimator slit on one side of the grating, and the telescope slit on the other, then the error introduced on collimation of the beam is approximately reduced to zero when

* 'Proc. Roy. Soc.,' A, vol. 132, p. 252 (1931).

† 'Astrophys. J.,' vol. 48, p. 125 (1918).

‡ 'Z. Physik,' vol. 61, p. 792 (1930).

the reflected parallel beam is converged on the telescope slit by the second mirror. Czerny showed experimentally that this was so, and that considerably improved images are obtained by using separate collimator and telescope mirrors. In fig. 1, M_4 and M_5 are the collimator and telescope mirrors respectively, and S_2 and S_3 their accompanying slits. The image of S_2 at S_3 obtained with the two-mirror system proved to be very good over even long

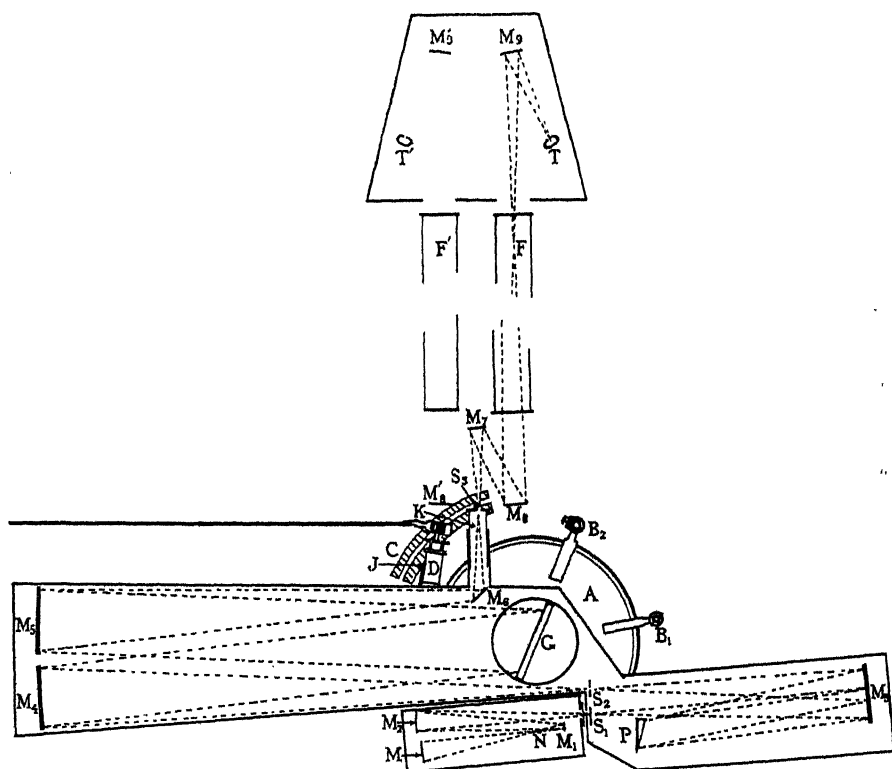


FIG. 1.

slit lengths, and the Hilger symmetrical slits were opened to their full aperture, 0.8 inch, instead of the usual 0.4-inch aperture.

The fore prism spectrometer has a focal length one half that of the grating spectrometer (16 inches and 32 inches respectively), so that the side of the prism need only be one-half the length of the rulings on the grating. We are indebted to Professor E. O. Salant, of New York University, for suggesting the focal lengths and mirror apertures used. The latter are of stainless steel, and of aperture 4 inches diameter; they were made by Zeiss.

The general design is straightforward; radiation from a Nernst filament, N,

is reflected in the mirror M_1 and converged on to the collimator slit of the fore prism spectrometer by M_2 . M serves as a backing mirror that increases the solid angle of radiation passing from the filament to the first slit of the spectrometer. It is the image of N in M_1 , instead of N itself, which is converged on the first slit; this device increases the available distance between N and M_2 , and facilitates orientation of the backing mirror to give a cone of radiation coincident with that falling on M_2 and converging on S_1 from N .

The radiation diverges from S_1 to M_3 , and is there collimated and reflected towards an 18° rocksalt prism, P . The back surface of this prism is platinum sputtered, and this mirror reflects the radiation through the prism to M_3 , which now converges it to S_2 , the collimator slit of the grating spectrometer. The prism was obtained from Messrs. Steeg & Reuter, of Bad Homburg, Germany; it is mounted on a theodolite circle readable to 30 seconds of arc. From S_2 the radiation diverges to M_4 , which collimates it and reflects it towards the grating, G . The only grating available was an echelette of 3600 lines per inch reflecting in the direction of the first order spectrum between 3.5 and 4μ ; it was ruled by R. W. Wood on a chromium-plated copper blank, the rulings being approximately 3 inches in length and extending over 4 inches. It is mounted on another theodolite circle readable to 1 second of arc; it is a standard circle of Messrs. E. R. Watts & Son, London, and we are much indebted to this firm for redesigning the mountings to suit this spectrometer. From the grating the parallel beam passes towards M_5 , and is converged towards M_6 , which reflects it to the telescope tubing which slides in a second tube firmly fixed to the spectrometer casing. This helps the focussing of the image of S_2 on S_3 .

Absorption Tubes and their Mirrors.

From S_3 the radiation diverges to the mirror M_7 , which is the mirror M_3 of the monochromator described in Part IV (*loc. cit.*). One important modification of the apparatus described in that paper was found necessary; the grating spectrometer has slits 0.8 inch long, as compared with slits of 0.4 inch used in the prism spectrometer; hence radiation diverging from the image F filled much more than the $1\frac{3}{4}$ -inch aperture of the end plates of the absorption tube; this difficulty was overcome by bringing the image, F , to a point near this end plate, and using a larger aperture mirror, M_9 , to converge the radiation on to the thermopile. The rapid divergence of the extreme rays is shown in fig. 2. O is an object illuminated by a cone of radiation whose solid angle is defined by the extreme rays intersecting a mirror, R , at A and B ; these converge

towards I as shown, one extreme ray crossing the axis at C, and the other passing directly to one end of I; the solid angle of radiation diverging from I is therefore defined by the lines drawn from C to either extremity of I, and the angle may be considerable. The position of I is accordingly adjusted so that the absorption tubes occupy a position KLMN relative to I.

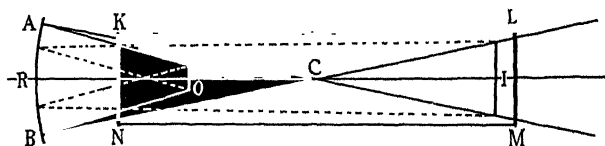


FIG. 2.

Spectrometer Casing and Mounting.

The fore prism spectrometer and grating spectrometer are mounted on a steel base $\frac{1}{4}$ inch thick, which is screwed to brick piers at either end and has four rigid supports near the centre. The brick piers are in turn screwed to a firm table on which the circles stand. The various mirrors and S_2 are in mounts screwed to the steel base; these mounts are provided with adjustments for rotation of the mirrors about vertical and horizontal axes, and S_2 may be rotated through a small angle about a horizontal axis passing through the centre of S_2 . The prism and grating tables have the usual adjustments. The axle supporting the grating spectrometer table passes through the centre of a rubber ring, making the joint air-tight, and avoiding the imposition of any strains on the circle due to uneven expansion of the spectrometer case, the bearing for the spectrometer table is screwed to the steel base plate; this again gives a roughly air-tight joint, and affords the necessary bearing for the axle of the theodolite circle; no great accuracy is required of this circle provided it has a reasonably good fine adjustment.

A pressed steel casing encloses the spectrometer, and carries the first slit S_1 ; the casing is screwed to the base plate, and angle brackets give it rigidity where the telescopic tube, T, is welded to it. A lid is fitted and gives an approximately air-tight joint. Nozzles are fitted to either end of the casing so that a stream of air freed from carbon dioxide and water vapour can be passed through the instrument.

Method of Use.

The grating table is fitted with a micrometer screw and drum, D, similar to those supplied with the Hilger direct reading spectroscopes. The drum is

uniformly graduated throughout the length of the spiral groove in which the index, J, moves. The end of this drum is fitted with a disc in which two slots are cut; two pins from a disc, rotated by worm motion from a distance, engage in the slots and so rotate the drum. It is read from a distance by means of a telescope. The disc with its two pins cannot remain fixed when the table is rotated through more than one degree, and this is allowed for by mounting the disc and its worm motion on a slide rest cut in a circle concentric with the axis of the table. The slide rest can be clamped in any position, with the pins passing through the slots in the end of D. Thus, when investigating a band the turntable is rotated to approximately the calculated position of the band and the pins are engaged in the slots of the drum; the band is then located by rotating D from a distance (actually at the galvanometer scale), and then finally carefully investigated by a rotation of D through the successive graduations on its surface. The band is plotted with readings of D as abscissæ, and the angles of required points are obtained by setting D at the corresponding graduations. The drum must, of course, be rotated in one direction when investigating the band, and in the same direction when determining the angles corresponding to given graduations on the drum, otherwise errors due to backlash of the micrometer screw may become considerable. This method of operating the spectrometer enables all the observations to be made by one person from one position in the room.

Setting-up the Spectrometer.

The mounts for the mirrors and slits were designed so that the centres of these parts were at one height above the base plate. The remaining adjustments were made with monochromatic light obtained by focussing a mercury arc on S_1 ; with M_3 approximately adjusted, the various mercury lines were observed near S_2 , and M_3 could then be accurately adjusted for parallel light, and the slit S_2 rendered parallel to the image of S_1 . M_4 was correctly focussed by reflecting light back towards it from the grating table by means of a plane mirror at the grating surface; M_5 was then set approximately at the correct position, and the final focussing of the image of S_2 on S_3 was made by the telescopic adjustment. An alternative to the use of monochromatic radiation is to focus the image of the Nernst filament on S_1 , and to replace P and G with plane mirrors; this gives more intense illumination and was used in the preliminary work. The mirrors M_5 and M_6 were actually adjusted by first viewing (from S_3) M_5 in M_6 , and then G in both M_5 and M_6 . A similar method was adopted for setting up mirrors M_7 to M_9 , as the visible light traversing this

system is rather weak. The mirrors were removed from the prism and grating tables, and the grating was rotated until the most intense red spectrum fell on S_3 ; looking along the axis of the absorption tube the observer can see M_7 in M_8 , and S_3 in the image of M_7 in M_8 ; finally M_9 was similarly adjusted and the thermopile placed at the image of S_3 formed by M_9 .

Procedure.

When the various components of the spectrometer have been correctly adjusted, there remains to be determined the correct setting of the prism and grating to obtain radiation of given wave-length traversing S_3 . The calibration of the prism need only be approximate, for so long as the wave-lengths 2 and 4 μ can be distinctly separated, the final setting of P can be made by means of the fine adjustment. Thus, if the grating be set so that wave-lengths of 4 μ in the first order spectrum traverse S_3 , then on rotating P, from the position where visible wave-lengths traverse S_2 towards longer wave-lengths, the observed galvanometer deflections pass through one maximum, corresponding to the wave-lengths near 2 μ in the second order, and then a second maximum appears due to the wave-lengths near 4 μ in the first order; the second maximum gives the correct setting of P. A maximum may appear of wave-length between 2 μ and the visible corresponding to a third order spectrum, but a circle readable to 30 seconds of arc is sufficiently accurate to avoid any confusion of the maxima, and the final setting of the prism can be made by means of the fine adjustment while observing the galvanometer deflection at each setting.

The grating can be set approximately by means of the different orders of visible spectra, but final settings and any observations of a band require calculation by means of the usual dispersion formula. Thus for a spectrum of the n th order we have

$$a(\sin \theta + \sin \phi) = n\lambda, \quad (1)$$

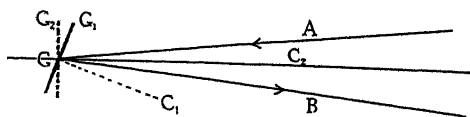


FIG. 3.

where a is the grating constant, θ is the angle of incidence, and ϕ is the angle of reflection. The grating as actually used is shown in fig. 3; AGC_1 is the angle θ , and BGC_1 the angle ϕ . The position of the grating when the zero

order appears in S_3 is GG_2 , and the normal occupies a position GC_2 . If ψ be the angle turned through by the grating from the zero order to the wave-length λ in the n th order, then ψ is equal to C_2GC_1 ; hence θ is $(\psi + \alpha)$, and ϕ is $(\psi - \alpha)$, where 2α is the angle between the incident and diffracted rays. Formula (1) therefore assumes the simple form

$$n\lambda = 2a \sin \psi \cos \alpha. \quad (2)$$

The grating accordingly is set by observing the zero order in S_3 , and rotating the turntable through an angle ψ determined from (2). Conversely, the wave-length corresponding to any setting of the turntable is determined from (2) on insertion of ψ .

Errors.

Errors in working the spectrometer arise from errors in ψ and α . Differentiating (2) for the first order spectrum, we have

$$\Delta\lambda = 2a (\cos \psi \cos \alpha \Delta\psi - \sin \psi \sin \alpha \Delta\alpha). \quad (3)$$

Since both ψ and α are small, error in ψ is the larger factor in the error, $\Delta\lambda$, occurring in λ . Determination of ψ involves reading two angles, that of the zero order spectrum and that of the table setting. The first point is the accuracy with which these readings can be made, and this involves the accuracy of the grating circle; this can be read to 1 second of arc, but the turntable as originally supplied had two micrometer eyepieces, one at each end of a diameter, and one of these had to be removed as it projected into the Nernst filament housing. The accuracy of the circle was tested with the two eyepieces in position, and readings on the two coincided to within 3 seconds of arc and we shall assume this to be the accuracy of the circle, and the error in setting the grating table. The greatest error arises in determining the position of the zero order spectrum; the position of S_3 renders impossible observations by an eyepiece, but a method which is quite as accurate is the following: the slit is covered with thin paper, and D is rotated until the edge of the image of S_2 is just visible; the corresponding setting of the circle is noted, and D is again rotated until the other edge of the image is visible, and the circle again read. The mean of these two readings then gives the position of the zero order. The method is as accurate as that involving an eyepiece, but observations to within much less than 7 seconds of arc appear impossible. Hence the total possible error in ψ is 10 seconds of arc; inserting this value in (3), and taking ψ as 17° and α as 5° , we obtain an error in λ equal to 0.0007μ .

The permissible error in α must be determined: inserting $\Delta\lambda = 0.0001 \mu$ in (3), we find $\Delta\alpha$ approximately 1 minute, and α must be known accurately to within this value; this is easily attained by rotating G so that the zero order falls firstly on S_2 , and secondly on S_3 .

Wave-lengths determined with this instrument should therefore be accurate to 0.001μ . Whilst the spectrometer was in use, a possible source of error was recognised; the reading of the zero order spectrum does not, unfortunately, remain constant over intervals of time of the order of 1 day, probably because of uneven expansion of the casing due to the heating lamps used to maintain the rocksalt parts at a slightly higher temperature than the surroundings; this involves rotation of the telescope tube carrying S_3 . The error can be eliminated by observing the position of the zero order spectrum before and after determining the contour of a band. Reproducible results were obtained by this method.

Resolving Power.

This is also determined from formula (3). Inserting 1 minute of arc for the value of $\Delta\psi$, we obtain $\Delta\lambda$ equal to 0.004μ ; but since the direction of incidence is constant, the diffracted wave-lengths must spread at the rate 2ψ , i.e., two diffracted wave-lengths separated by an angle of 1 minute must differ in wave-length by 0.002μ . Again, a slit width of 0.01 second subtends an angle of approximately 0.6 minute at the grating; hence slit widths of 0.01 second, or 10 divisions, contain a wave-length range of 12 Å., or 0.8 wave-numbers, near 4μ . Between 3.5 and 4μ , galvanometer deflections of 1.5 cm. were obtained with this slit width, and if the variations in percentage absorption are marked, this deflection is large enough to obtain reliable observations.

The bands actually observed all lay beyond 4μ , and the least slit widths used were 20 divisions, or 24 Å for the 4.003μ band of SO_2 . Other bands near 4.35 and 4.6μ required 40 divisions or approximately 50 Å.

Elimination of Water Vapour and CO_2 .

The spectrometer was ultimately to be used over regions where carbon dioxide and water vapour have considerable absorptive powers, and the instrument was therefore designed to be roughly air-tight. The method adopted to eliminate these impurities consisted in displacing the air in the case by dry purified air, and maintaining a slight excess pressure inside by means of a stream of gas from evaporating liquid air in a Dewar vessel. As the longer wave-length band of SO_2 overlaps the 4.25μ band of CO_2 , the further pre-

caution was taken of covering the slits S_1 and S_3 with rocksalt windows, and placing sticks of caustic soda in the case. When the contents of the small Dewar vessel were allowed to evaporate overnight, it was found that the CO_2 absorption was reduced approximately to that observed with the prism spectrometer; this residual absorption was presumably due to absorption in the path outside the casing; we hope eventually to enclose the total path length. When the spectrometer as treated was left over the week-end without liquid air evaporating into it, no appreciable increase in the absorption due to CO_2 was observed.

Before the impure air was removed from the case, it was not possible to obtain measurable deflections in the neighbourhood of the maximum of the CO_2 band at 4.25μ ; after purification readings were obtained throughout this region, and the setting of the spectrometer was checked against Barker's values for the band in question.*

Our acknowledgments are due to Professor F. G. Donnan, C.B.E., F.R.S., for his continued interest and encouragement; to Imperial Chemical Industries, Ltd., from whose grant to these laboratories certain portions of the apparatus were purchased; to Mr. G. W. Alliss, of this Department, who designed and made many of the parts; and to the Department of Scientific and Industrial Research for a Senior Award to A. B. D. C.

Summary.

- (1) The construction of an infra-red grating spectrometer for use as a double monochromator is described.
- (2) The method of working such a spectrometer is fully explained.
- (3) The experimental errors and the resolving power are calculated.

* 'Astrophys. J.,' vol. 55, p. 391 (1922).

The Resonance Spectrum of Hydrogen.

By K. R. RAO, D.Sc., Science College, Andhra University, Waltair, India, and
J. S. BADAMI, Ph.D., Physikalisches Technische Reichsanstalt, Berlin.

(Communicated by A. Fowler, F.R.S.—Received July 15, 1932.)

[PLATE 14.]

Introductory.

The success of the hollow cathode discharge in helium in exciting the spectra of various metals suggested that a similar method might be applied to excite the spectrum of arsenic. It was in the course of this work that plates were obtained which revealed with remarkable intensity the Lyman series of hydrogen, extending down to the 15th member. The most interesting feature of the series, however, is the anomalous distribution of intensity among its members. Instead of the intensity gradually falling off as we go to higher members, a sudden enhancement is observed in the intensity of the 10th and 11th members and then an abrupt fall so that the remaining members are only just observable. The appearance of the series is such as immediately to suggest the occurrence of a kind of resonance effect. The experiments were therefore repeated under various conditions in order to investigate the cause of this observed anomalous intensity and to discover, if possible, the type of impact involved in the production of such a phenomenon. In two preliminary notes,* the main conclusions which have been obtained have been reported and it is the purpose of this paper to describe these investigations and the conclusions reached.

Experimental.

The experimental use of the hollow cathode source is well known and has been described previously by various writers† but on account of the importance of the phenomenon that has been observed, it is considered worth while to present details of the method employed. Some metals, like aluminium, are generally used directly as the cathode, but Paschen‡ has found that the source can be adapted also for metals of low melting points or those which vaporize

* 'Nature,' vol. 128, p. 585 (1931), and vol. 129, p. 869 (1932).

† Paschen, 'Ann. Physik,' vol. 71, p. 537 (1923), and Sawyer and Paschen, 'Ann. Physik,' vol. 84, p. 1 (1927).

‡ Paschen, 'SitzBer. preuss. Akad. Wiss. Berl.,' vol. 29, p. 207 (1927).

rapidly, in which case the metal in question may be placed in a carbon (or tungsten) hollow cathode. The cathode in these experiments, was a thin hollow cylinder of carbon, C, fig. 1, of length 3.5 cm. and diameter about 1.5 cm. The cylinder was closed at one end N, which terminated in a kind of box of length 0.5 cm., the inner wall W of which was perforated. This box served as a reservoir to contain metallic arsenic, whose vapour entered the cathode

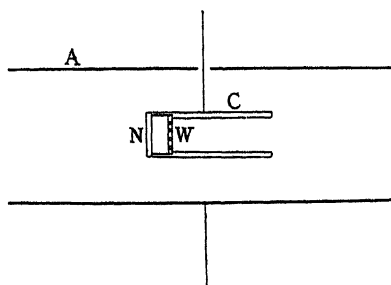


FIG. 1.—Hollow cathode source.

through the perforated wall. A similar cathode was used by Takahashi* in his experiments on Zn and Cd.

The anode A, made of thin sheet steel, was a co-axial cylinder of length 13 cm. surrounding the cathode and just fitting into the discharge tube, which consisted of a double-walled quartz tube having an overall length of 35 cm. and internal diameter of about 4 cm. One side (that which was away from the closed end of the cathode) of the quartz tube was ground and fitted exactly over the ground cone of the vacuum spectrograph carrying the slit, while the other end was closed with a plate glass or quartz window. A diagram of all the parts of the apparatus is shown in fig. 2.

R is a helium reservoir; G is a 4-litre globe into which any desired quantity of the gas may be admitted for circulation through the apparatus. A, B, C are liquid-air cooled traps, C containing charcoal. These served to absorb Hg vapour and other impurities. The tube D containing copper oxide could, if necessary, be heated to absorb hydrogen. To start with, the whole system was exhausted to a hard vacuum; helium was then admitted into the globe G, the pressure in it being registered by the manometer M. The mercury vapour pump P was then set in action and it kept the admitted quantity of helium in circulation in the whole system, including the spectrograph. This process of circulation was continued throughout the experiment; exposures were begun only after the impurities were all absorbed, which took some hours,

* 'Ann. Physik,' vol. 3, p. 27 (1929).

particularly with a freshly made cathode of carbon. Such a cathode was also preliminarily treated with NaOH, washed and heated to bright redness before it was inserted into the discharge tube. Traces of neon contained in helium, always appeared as an impurity. A constant D.C. potential of 800 volts was applied to the electrodes and by a suitable resistance a current of about 250 m. amps. was maintained throughout the exposure. The spectrum was photographed by a vacuum spectrograph of 1 metre radius, giving a

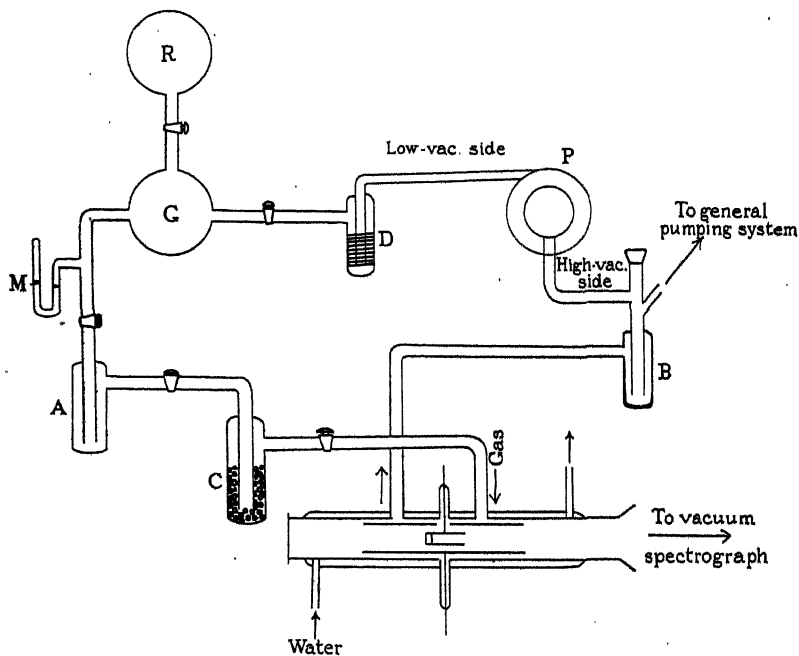


FIG. 2.

dispersion of about 17 Å. per millimetre. Exposures varying from 6 to 8 hours were found necessary and the entire region of the spectrum between λ 2800 and λ 500 was obtained on a single plate.

It should be stated that in all these experiments hydrogen was not introduced into the tube from any external source, but that the faint traces of the gas, present as an impurity in the helium supply and in the various parts of the apparatus, served to exhibit the phenomena described here. The tube of copper oxide which was intended to absorb the hydrogen, was not heated, except in those experiments where a strong spectrum of arsenic was mainly required. When the container in the cathode was once filled with arsenic metal, the metal lasted for about five or six exposures but the relative intensity

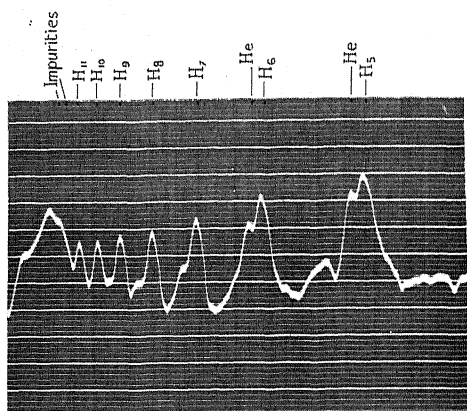
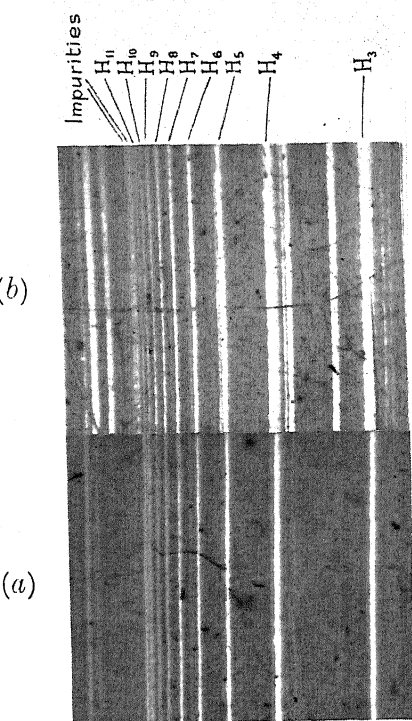


FIG. 4b.

FIG. 3.—Lyman series of hydrogen in a hollow cathode in helium. (a) Shows intensity anomaly.

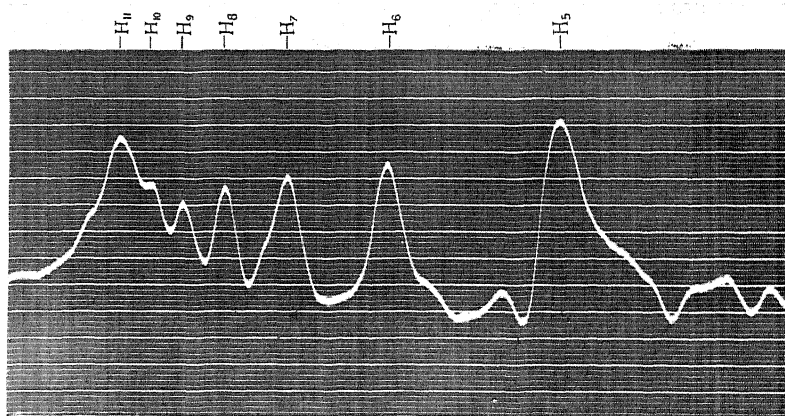


FIG. 4a.

FIG. 4.—Microphotometer curves of the Lyman hydrogen series.

(a) Shows the anomalous intensity.

(b) Shows the normal intensity.

of the resonance lines of As I, λ 1972, λ 1937, and λ 1890 due to the combination $4p\ ^4S-5s\ ^4P$, gradually diminished with successive exposures till, when all the arsenic had evaporated (or had been sublimed) these lines were absent altogether. The relative strength of these fundamental lines were used as a rough indication of the pressure of arsenic in the tube. For varying the pressure of helium, different quantities of the gas were admitted into the globe G for circulation. No attempts were made at measuring the exact pressure of the gas in the discharge tube but the difference of pressure indicated by the manometer M, during the progress of the experiment, was taken as an index. This pressure was varied from 20 mm. to 2 mm. in the experiments. Plates were taken of the discharge, in the absence of arsenic, with varying pressures of helium and also with various amounts of arsenic in the tube, while admitting the same amount of helium into the globe.

A few experiments were also made with neon instead of helium in the discharge tube, the results obtained with this gas being identical with those for helium.

Results and Discussion.

Two very significant results emerged from these experiments.

(1) Only the first three members of the Lyman series of hydrogen had previously been photographed. However, the authors have been informed, that on various occasions, while experimenting with this source, Professor Paschen and other workers in his laboratory have observed, although extremely faintly, a few more members. In this work the series has been photographed, for the first time, as far as the 15th member. All other conditions of the discharge being maintained the same, an increase in the pressure of helium alone (up to the maximum employed) was found to favour the excitation of the higher members of the series. The absence of the ultra-violet secondary spectrum of hydrogen on any of these plates is worth noting. It is plausible to assume that transfer of energy might have taken place through collisions of the second kind frequently occurring between the atoms of helium in the metastable state and the hydrogen molecules, resulting in an abundance of hydrogen in the atomic state. This would account for the excitation of the primary and the suppression of the secondary spectrum of hydrogen.

It may be interesting to recall here the investigations made to extend the Balmer series of hydrogen. Besides Wood's classical method* of using long

* 'Proc. Roy. Soc.,' A, vol. 97, p. 455 (1921).

vacuum tubes and of Herzberg's method* of sending an electrodeless discharge of high frequency through hydrogen at suitable pressure, Whiddington† has reported a method of photographing the Balmer series as far as the 20th member by utilising a hydrogen tube having a hot cathode of tungsten wire; the significant observation was made that near the cathode, within the dark space, the secondary spectrum was absent entirely, only the primary Balmer series appearing. The phenomenon might be due, as it was suggested, to direct electronic bombardment, as the space near the cathode is rich in electrons, a condition closely akin to that resulting in the emission of the complete Balmer series in stellar spectra. The fact that Balmer lines are better developed when the hydrogen is mixed with rare gases seems also to have long been known. Merton and Nicholson‡ obtained to H_{14} , by using a trace of hydrogen mixed with helium at about 41 mm. of pressure. Of particular importance is the observation made by Takamine and Suga§ of the Balmer series down to H_{25} in a neon discharge tube in which hydrogen was present only as a trace of an impurity. The production of the series to such a high member has been ascribed to the presence of hydrogen very richly in the atomic state, formed by second collisions with metastable neon atoms. With argon, the series was not so well developed but observed only to H_{16} . These results are obviously similar to those observed in the present experiments.

The following table gives the wave-lengths of the lines of the Lyman series, as directly measured on the present plates. The first three lines, λ 1215, λ 1025, λ 972, are too broad for accurate measurement; the eight lines, H_4 to H_{11} , have been measured in the second order by interpolation between the first order carbon line λ 1930.94 and the third order helium line λ 584.38; and the last four lines, H_{12} to H_{15} are first order measures. Owing to the faintness of these latter, the error in the wave-lengths is likely to be large.

Line.	Wave-length.	Line.	Wave-length.	Line.	Wave-length.
H_4	949.74	H_8	923.17	H_{12}	917.16
H_5	937.81	H_9	920.99	H_{13}	916.42
H_6	930.76	H_{10}	919.38	H_{14}	915.80
H_7	926.24	H_{11}	918.12	H_{15}	915.30

* 'Ann. Physik,' vol. 84, p. 553 (1927).

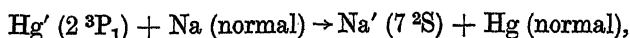
† 'Phil. Mag.,' vol. 46, p. 604 (1923).

‡ 'Phil. Trans.,' vol. 217, p. 237 (1916).

§ 'Sci. Pap. Inst. Phys., Chem. Res. Tokyo,' vol. 14, p. 117 (1930).

(2) The second and more important result that has been obtained is the anomalous intensity exhibited by the members of the series. Fig. 3 (a), Plate 14, is a photograph of the discharge with arsenic in the cathode and it is clearly seen that there is a considerable enhancement in the intensity of the 10th and the 11th members. After the 11th line the intensity diminishes so rapidly that the next four members could be traced only with difficulty on the negative and are not capable of reproduction. Fig. 3 (b), Plate 14, is a picture of the discharge under the same conditions except that arsenic was not present. The anomaly disappears altogether; all the above members are distinctly seen, with the intensity diminishing gradually as we proceed to the higher members, as in the case of a normal series. With a cathode freshly filled up with the metal, the enhancement of the 10th and 11th lines is very marked; the intensity of these lines approaches that of even the third or second member of the series. Figs. 4 (a) and (b) Plate 14 are microphotometer curves taken of the plates reproduced in figs. 3 (a) and (b) respectively, indicating the relative intensity of the members in each of these two cases and the anomalous feature is obvious from the first curve. It is therefore inferred that this phenomenon occurs only in the presence of arsenic, becoming more prominent with greater pressure of arsenic vapour in the discharge tube. Care has been taken to establish that there are no arsenic or associated impurity lines that could account for the observed anomalous intensities of the two lines in question.

Such phenomena of enhanced intensities of certain lines in similar discharges have often been mentioned. Beutler and Josephy* showed that when a mixture of mercury and sodium vapours was excited by $\text{Hg } \lambda 2537$, a maximum of excitation was observed for the Na doublet at $\lambda 4420$ corresponding to (2P—7S), whose excitation energy agrees most closely with that of the state 2^3P_1 of mercury. Addition of a foreign gas caused the appearance of a second maximum at 2P—5S of Na, corresponding to the energy of the metastable state 2^3P_0 of mercury. These same maxima of intensity were observed also by Webb and Wang† who, by controlling the experimental conditions, so arranged that collisions of the atoms of sodium took place more effectively with excited mercury atoms than with mercury ions. The processes occurring in these experiments may be represented by the reaction equation



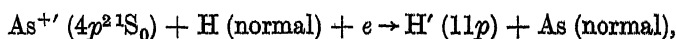
in which the dash against the symbol represents that the atom is excited to

* Beutler and Josephy, 'Naturwiss.', vol. 15, p. 540 (1927).

† 'Phys. Rev.', vol. 33, p. 329 (1929).

a state, the designation of which is indicated within the brackets. It must be noted that, in these cases, a "close resonance" is of great importance for the transfer of energy between the colliding partners to take place with the greatest probability.

It is considered that the present phenomenon is analogous to these observations of the "sensitised fluorescence" of sodium vapour, the requisite energy for the enhanced members of the Lyman series being transferred to the hydrogen atom, by impacts of the second kind, from the atoms of arsenic. It might be suggested that, instead of the atomic states of arsenic being involved in the transfer, the molecular energy states of some compound of As and H are responsible for the production of the phenomenon. Apart from the indefiniteness of this supposition, the possibility of it seems to be precluded, as there is no evidence on the plates of any trace of bands characteristic of As-H. It may be worth while mentioning that besides the hydrogen series, the arc spectrum of carbon was also strongly developed; but, so far as the writers could judge, no intensity anomaly was observable in the lines of this element. It is thought that in the production of the phenomenon, the following reaction had taken place,



in which $\text{As}^{+'}$ represents an ionised atom of arsenic in an excited state, this excited state being here supposed, on account of the considerations to be presently put forward, to be the metastable state $4p^2 {}^1\text{S}_0$ which is indicated in brackets; " e " represents an electron involved in the collision. The formation of these ionised and excited atoms of arsenic might take place by one of many processes. The excitation of metallic spectra by discharges in a hollow cathode in a rare gas atmosphere was discussed by Paschen,* who was the first to use it as a spectroscopic source, and also by Frerichs,† Takahashi‡ and Sawyer.§ There is not entire agreement as to the exact nature of the processes involved, *i.e.*, whether the collisions of the excited or ionised rare gas atoms take place with neutral or with excited or ionised atoms of the metal is not quite clear. As Sawyer has concluded, any of these collisions may occur under suitable conditions. The process would be complicated in the case of molecular gases which involve the additional factor of the dissociation energy to be taken into consideration. However, we are not much concerned with the exact type of

* 'SitzBer. preuss. Akad. Wiss.,' vol. 29, p. 207 (1927).

† 'Ann. Physik,' vol. 85, p. 362 (1928).

‡ 'Ann. Physik,' vol. 3, p. 49 (1928).

§ 'Phys. Rev.,' vol. 36, p. 44 (1930).

these impacts, but only with the final result; namely, the formation of the atoms of the metallic vapour in an ionised and excited state, regarding which there is general agreement. The available energy for the excitation of the metallic atoms is limited only by the energy of ionisation of the rare gas atoms. Assuming then the presence, in the hollow cathode source, of a large number of atoms of arsenic in the metastable state $4p^2\ ^1S_0$ of As II, the reaction stated above is capable of explaining the observed phenomenon of the enhancement of the intensity of the 11th member of the Lyman series. That there also really exists a close resonance between the colliding atoms, which is of importance for the transference of energy between the two and which makes the reaction to proceed with the greatest probability, is evident from the following equality,

	volts.	ν (cm. ⁻¹).
(A) Excitation energy of the 11th member of the Lyman series of hydrogen	13.44	\equiv 108916
(B) First ionisation potential of arsenic	10.5	\equiv 85000
(C) Energy required to excite the ionised atom of arsenic from the normal state 3P_0 to the metastable state 1S_0	2.79	\equiv 22599

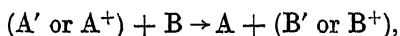
The data for the items (B) and (C) are taken from two papers by Mr. A. S. Rao which are in the course of publication.* A discussion regarding the value (B) of the first ionisation potential of arsenic, in which the suggested probable error is 2000 cm.⁻¹, is given in one of the above papers. It is clear that the excitation energy (A) is equal to (B) + (C), within the limits of error involved in the estimate of (B).

The explanation of the phenomenon is then, that the arsenic atoms in the metastable state $4p^2\ ^1S_0$ of As II, in returning to the normal state of As I, transfers this *total energy of excitation and ionisation* to the hydrogen atoms by collisions of the second kind. This would also explain the necessity, in these experiments, of absorbing all the hydrogen in order to photograph the spectrum of arsenic intensely. That the phenomenon cannot be explained as due either to direct electronic impacts or to impacts of the second with only the excited rare gas atoms, is evident from the observed experimental fact of the necessity of the presence of arsenic for the occurrence of the phenomenon.

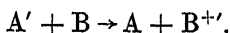
* 'Proc. Phys. Soc.,' vol. 44, p. 343 (1932); 'Proc. Phys. Soc.,' (*in course of publication*).

A New Type of Collision.

If the above interpretation is accepted, we meet here with a type of collision which, so far as the authors are aware, has not been observed in any of the previous experiments in this field. Since Franck* has generalised the theoretical conclusions of Klein and Rosseland,† a large number of investigations have been carried out demonstrating energy transfers in a variety of processes between electrons, atoms and molecules.‡ We have ample evidence of a direct electronic impact producing excitation or ionisation, or simultaneous ionisation and excitation in atoms and molecules. And these methods have led to the direct experimental determination of the critical potentials of atoms and molecules. There is yet another class of experiments dealing with the so-called collisions of the second kind in which this excitation or ionisation is produced by transference of the requisite energy from an excited or ionised atom. The reaction in these cases, may be represented generally by the equation

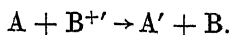


in which A' denotes the excited and A^+ , the ionised atom A . Of particular interest are the experiments of Duffendack and Smith,§ who demonstrated the production of *simultaneous ionisation and excitation* in molecules by impacts with excited rare gas atoms, the equation of the reaction being,



And the usual excitation of metallic spectra in a hollow cathode source, indicates the occurrence of this process in the case of collisions between two atoms.

The present experiments, on the other hand, demonstrate the occurrence of a type of collision in which the reverse process is involved, which is represented thus,



The writers believe that this is the first time that the occurrence of such a reverse process has been demonstrated, apart from its importance in being associated with the excitation of the most elementary of all known spectra, namely, the Lyman series of hydrogen.

* 'Z. Physik,' vol. 9, p. 259 (1922).

† 'Z. Physik,' vol. 4, p. 46 (1921).

‡ Franck and Jordan, "Anregung von Quantensprünge durch Stosse" (Julius Springer).

§ 'Phys. Rev.,' vol. 34, p. 68 (1929).

In the light of these experiments, it may be suggested that it would be interesting to see if a similar phenomenon could be observed with antimony or bismuth; the intensity anomaly might be shifted in either of these cases to other members of the Lyman series. Experiments should also be possible with other properly chosen elements, in which the enhanced intensity might occur among the members of the Balmer series.

We take this opportunity of expressing our thanks to Professor M. N. Saha, F.R.S., for his valuable advice and suggestions in the preparation of the paper. The experiments were performed at the Physikalisches-Technische Reichsanstalt, and it is a pleasure to acknowledge our gratitude to Professor F. Paschen for helpful advice and several discussions during their progress. To Dr. A. L. Narayan we are indebted for the curves taken on the Cambridge microphotometer recently set up by him at the Solar Physics Observatory, Kodaikanal.

Summary.

In the course of experiments on the spectrum of arsenic with the Paschen hollow cathode discharge, photographs have been obtained showing the Lyman series of Hydrogen very strongly developed; the series is traceable down to the 15th member. The peculiar feature has been noticed that the intensity, instead of diminishing with successive members as in the case of a normal series, is enhanced markedly at the 10th and the 11th members, and then falls off abruptly, indicating the occurrence of a kind of resonance phenomenon. Further experiments have shown that this anomalous feature disappears in the absence of arsenic in the discharge tube but appears more markedly with increasing pressure of arsenic vapour. The explanation is suggested that the phenomenon is caused by a transfer of energy by collisions of the second kind between the arsenic atoms in the metastable state $4p^2\ ^1S_0$ of As II and the atoms of hydrogen. This explanation is supported by the equality of the excitation energies involved. A new type of impact is thus demonstrated in the occurrence of this phenomenon, for the *energies of ionisation and excitation of one atom are simultaneously transferred to another atom* by a single act of collision.

*The Calculation of the Terms of the Optical Spectrum of an Atom
with one Series Electron.*

By J. McDougall, Pembroke College, Cambridge.

(Communicated by R. H. Fowler, F.R.S.—Received July 19, 1932.)

1. The theoretical determination of the energies of the stationary states of an atomic system is bound up with the solution of the many-body problem—in particular, with the determination of wave functions of many-electron atoms. An exact solution is not known, but approximations to it have been made by Hartree,* Slater,† Fock‡ and Lennard-Jones.§ The method adopted is to replace the physical problem by an artificial one which admits of a solution, *e.g.*, Hartree replaces the actual many-body problem by a one-body problem with a central field for each electron. Generally, the Schrödinger equation for an atom of nuclear charge N is

$$\left\{ \sum_{i=1}^N \left(-\frac{1}{2} \nabla_i^2 - \frac{N}{r_i} \right) + \sum_{i>j=1}^N \frac{1}{r_{ij}} - E \right\} \Psi = 0, \quad (1.1)$$

using atomic units|| and the usual notation. The artificial system replacing (1.1) has the equation

$$\left\{ \sum_{i=1}^N \left(-\frac{1}{2} \nabla_i^2 - v_i \right) - E \right\} \Psi = 0, \quad (1.2)$$

v_i being a function of the co-ordinates of the i th electron only. Equation (1.2) is separable, and reduces to equations of the type

$$\left\{ -\frac{1}{2} \nabla_i^2 - v_i - E_i \right\} \psi = 0, \quad (1.3)$$

in the space co-ordinates of the i th electron alone. If the solutions of equations (1.3) are $\psi(\alpha|1)$, ..., $\psi(\pi|p)$, where the Greek letter is the label of the wave function, and the numeral or Roman letter indicates the electron whose co-ordinates are substituted, then a solution of (1.2) is

$$\Psi = \psi(\alpha|1) \psi(\beta|2) \dots \psi(\pi|p). \quad (1.4)$$

* 'Proc. Camb. Phil. Soc.,' vol. 24, p. 89 (1928).

† 'Phys. Rev.,' vol. 34, p. 1293 (1929).

‡ 'Z. Physik,' vol. 61, p. 126 (1930).

§ 'Proc. Camb. Phil. Soc.,' vol. 27, p. 469 (1931).

|| Unit of length, the first Bohr hydrogen radius; unit of charge, the magnitude of the charge on the electron; unit of mass, that of the electron; unit of action, $\hbar/2\pi$; unit of energy, $2\hbar cR$, *i.e.*, twice the ionisation energy of the hydrogen atom with fixed nucleus; unit of time, $1/4\pi cR$.

The form of wave function which must be assumed in order to satisfy Pauli's Exclusion Principle and be antisymmetric in the co-ordinates of all pairs of electrons, is the determinantal form

$$\Psi = \begin{vmatrix} \psi(\alpha|1) & \psi(\alpha|2) & . & . & . & \psi(\alpha|p) \\ \psi(\beta|1) & \psi(\beta|2) & . & . & . & \psi(\beta|p) \\ . & . & . & . & . & . \\ \psi(\pi|1) & \psi(\pi|2) & . & . & . & \psi(\pi|p) \end{vmatrix}, \quad (1.5)$$

which is the sum of the expressions obtained by permuting the co-ordinates 1, 2, ..., p in the product (1.4) and taking account of the signs of the permutations. Thus we obtain an approximate wave function for the whole atom in terms of the one-electron wave functions.

Hartree's method for determining energies of optical terms will be outlined, and then Slater's approximation to the many-body problem will be used to obtain more accurate expressions for these terms, and the equations will be solved by numerical methods.

2. Hartree's Method.

In atomic units, the wave equation for a stationary state of an electron with total energy E , in a field in which its potential energy is V , is

$$\nabla^2\psi + 2(E - V)\psi = 0. \quad (2.1)$$

For an attractive field V is negative, and E is negative for optical and X-ray terms; so it is convenient to write $V = -v$, and $E = -\frac{1}{2}\varepsilon$; ε is then the energy expressed as a multiple of the ionisation energy of the hydrogen atom. The relation between wave number and energy is now

$$\varepsilon = v/R, \quad (2.2)$$

and the wave equation becomes

$$\nabla^2\psi + (2v - \varepsilon)\psi = 0. \quad (2.3)$$

When the field is spherically symmetrical, i.e., $v = v(r)$, a solution of the type

$$\psi = R(r) \cdot S(\theta, \phi), \quad (2.4)$$

may be found, S being a spherical harmonic. The differential equation for the radial part of ψ is simplified by writing

$$P = rR, \quad (2.5)$$

when the equation reduces to the form

$$P'' + \left\{ 2v - \epsilon - \frac{l(l+1)}{r^2} \right\} P = 0, \quad (2.6)$$

which is more convenient for numerical work, as there is no term in P' . l is the order of the spherical harmonic, and on Schrödinger's interpretation, P^2 is the radial density of charge, when P is normalised.

(a) *The Self-consistent Field.*—The self-consistent field for an atom or ion is such that the wave functions for the electrons in that field give a distribution of charge, which, with the nucleus, reproduces the field. An initial field v is assumed, and the equation (2.6) is solved for each core electron (n, l). This equation has one solution which is zero at the origin, and one which is zero at infinity, for each value of ϵ ; but for certain values of ϵ these are the same and the solutions of the equation found by integrating outwards from the origin, and inwards from large r , will join smoothly at intermediate radii. The approximate characteristic ϵ value is found by trial and error methods, and the corresponding charge density is found from the wave functions P , the solutions of (2.6) with this value of ϵ . The sum of the charge distributions for all electrons but one is found, and the potential of the resulting charge is calculated, and compared with the assumed field for the one electron. Further approximations are made until the assumed and calculated fields agree sufficiently well. Then the sum of the density distributions of all the electrons is taken to be the real density distribution in the atom, and the corresponding potential gives the "self-consistent" field. The self-consistent field so found is characteristic of the particular atom in the particular state of ionisation considered, and involves no arbitrary functions or constants. The characteristic numbers ϵ of the various central field problems lie close to the energy values of the corresponding terms of the spectrum.

The self-consistent field is usually expressed in terms of Z , Z_p . Z , the "effective nuclear charge," is the point charge which, placed at the nucleus would give the same *field* at radius r as the actual field. Z_p is the "effective nuclear charge for potential," i.e., the point charge which, placed at the nucleus, would give the same *potential* at radius r as that of the actual field. These are related by the equations.

$$\left. \begin{aligned} Z_p &= -r \int_r^\infty Zr^{-2} dr \\ \frac{dZ_p}{dr} &= -\frac{Z - Z_p}{r} \end{aligned} \right\}. \quad (2.7)$$

$[Z]_{n,l}$ and $[Z_p]_{n,l}$ may be defined as the contributions to Z , Z_p from the spherically averaged distribution of charge of one n, l electron and one unit of positive charge on the nucleus, and then

$$[Z]_{n,l} = \int_r^\infty P^2 dr, \quad (2.8)$$

if P is normalised. $[Z]_{n,l}$ and $[Z_p]_{n,l}$ are related by equations similar to (2.7). The self-consistent field for Si^{+4} has been calculated, and is given in Table I. Approximations were made until the initial and final values of $2(2l+1)[Z]_{n,l}$, the contribution to Z from the whole n, l group, agreed within 0.001 for each group. The table* gives the normalised core wave functions, the contributions to Z from the $1s, 2s, 2p$ groups, and the Z_p values for the Si^{+4} ion.

Table I.—Self-Consistent Field for Si^{+4} .(a) Normalised Core Wave Functions, $P(\alpha|r)^\dagger$

r .	$P(1s r)$.	$P(2s r)$.	$P(2p r)$.	r .	$P(1s r)$.	$P(2s r)$.	$P(2p r)$.
0.00	0	0	0	0.6	0.021*	-1.246*	1.115
0.01	0.886*	0.229	0.007*	0.7	0.007	-1.083	0.968
0.02	1.542*	0.396	0.028*	0.8	0.002	-0.905	0.819
0.03	2.013	0.511*	0.060*	0.9	0.000*	-0.736	0.680
0.04	2.336*	0.584	0.100*	1.0	0	-0.586*	0.557
0.06	2.661*	0.629*	0.198*	1.2		-0.356*	0.361*
0.08	2.698	0.579	0.311	1.4		-0.209	0.227
0.10	2.566	0.467	0.429	1.6		-0.118*	0.139
0.12	2.347	0.317	0.547	1.8		-0.065*	0.083*
0.14	2.088	0.147	0.660	2.0		-0.035*	0.049*
0.16	1.823	-0.031	0.766*	2.2		-0.019	0.029
0.18	1.567	-0.208	0.864	2.4		-0.010	0.017
0.20	1.331	-0.377*	0.950*	2.6		-0.005*	0.009*
				2.8		-0.003	0.005*
0.25	0.854	-0.748	1.122*	3.0		-0.001*	0.003
0.30	0.528	-1.024*	1.231	3.2		-0.001	0.001*
0.35	0.319	-1.209	1.286	3.4		-0.000*	0.001
0.40	0.189	-1.313*	1.300	3.6		0	0.000*
				3.8			0.000*
0.5	0.064*	-1.349*	1.240	4.0			0

(b) *Hartree's Method for Optical Terms*.—We now have to evaluate the wave functions and ϵ numbers for the series electrons. Strictly, we should

* Throughout the tables given in this paper, the dot signifies that the digit following the last digit printed lies between 3 and 7, e.g., 14.13^* is a number lying between 14.133 and 14.137 . This makes for neater printing than the use of such an alternative form as $14.13\frac{1}{2}$.

† The core wave functions are drawn in fig. 1, α .

(b) Contributions to the Effective Nuclear Charge Z from the $1s$, $2s$, $2p$ groups, and the Effective Nuclear Charge for Potential, Z_p , for the Ion.

r .	Contributions to Z .			Z_p for ion.	r .	Contributions to Z .			Z_p for ion.
	$1s$ group.	$2s$ group.	$2p$ group.			$1s$ group.	$2s$ group.	$2p$ group.	
0.00	2.000	2.000	6.000	14.000	0.35	0.008*	1.692*	4.437*	6.409
0.01	1.994*	1.999*	6.000	13.535	0.40	0.002	1.531*	3.933	5.922
0.02	1.963	1.997*	6.000	13.086*					
0.03	1.899	1.993*	6.000	12.663	0.5	0	1.169	2.954	5.192*
0.04	1.803	1.987	6.000	12.268	0.6		0.827*	2.114*	4.721
					0.7		0.554	1.462	4.427
0.06	1.547*	1.972*	5.997	11.562	0.8		0.355*	0.981*	4.249
0.08	1.257	1.957	5.989	10.956*	0.9		0.221	0.644*	4.143*
0.10	0.977*	1.946	5.973*	10.431	1.0		0.133*	0.415	4.081*
0.12	0.735	1.940	5.945	9.964					
0.14	0.537*	1.937	5.901	9.541*	1.2		0.046	0.164	4.026
0.16	0.384*	1.937	5.840	9.151*	1.4		0.015	0.061*	4.008
0.18	0.269	1.936	5.758*	8.787	1.6		0.004*	0.022	4.002
0.20	0.185*	1.933	5.661	8.444	1.8		0.001*	0.008	4.000*
					2.0		0	0.002*	4.000
0.25	0.067	1.899*	5.333*	7.665*	2.2			0.001	4.000
0.30	0.020*	1.818*	4.915*	6.989	2.4			0	4.000

ϵ values $1s$ 141.2
 $2s$ 14.13*
 $2p$ 11.23

$-2v_0 = 93.60$.

v_0 is the potential at the nucleus of the negative distribution of charge due to the outer electrons.

work out a self-consistent field problem for each state $3s$, $4p$, etc., of the series electron, and so obtain the wave functions and ϵ values corresponding to that state by the method indicated above, but this would be too laborious. The self-consistent field of the whole atom with the series electron is not calculated, but good approximations to the wave functions of the series electron are obtained by solving the equation (2.6) for each state of the series electron, in the field of the *unperturbed* core. Thus, for the optical terms of Si^{+3} , instead of carrying out separate self-consistent field calculations for the configurations $(1s)^2(2s)^2(2p)^6(3s)$, $(1s)^2(2s)^2(2p)^6(4p)$, etc., we solve the equations for the electrons $3s$, $4p$, etc., in the undisturbed field of the core $(1s)^2(2s)^2(2p)^6$, i.e., in the field of the ion Si^{+4} . The appropriate l values for the series electron are substituted ($l = 0$ for s , $l = 1$ for p terms, etc.), and the ϵ values for which the solutions of the equation (2.6): (i) zero at the origin, (ii) zero at infinity, join smoothly at intermediate radii, are found. The corresponding P functions are the wave functions of the central field problems used in section 4; some of these are drawn in figs. 1, b and 1, c . Fig. 1, b , gives wave functions for the range of nuclear distance $r = 0$ to $r = 4$, which is the effective range of integra-

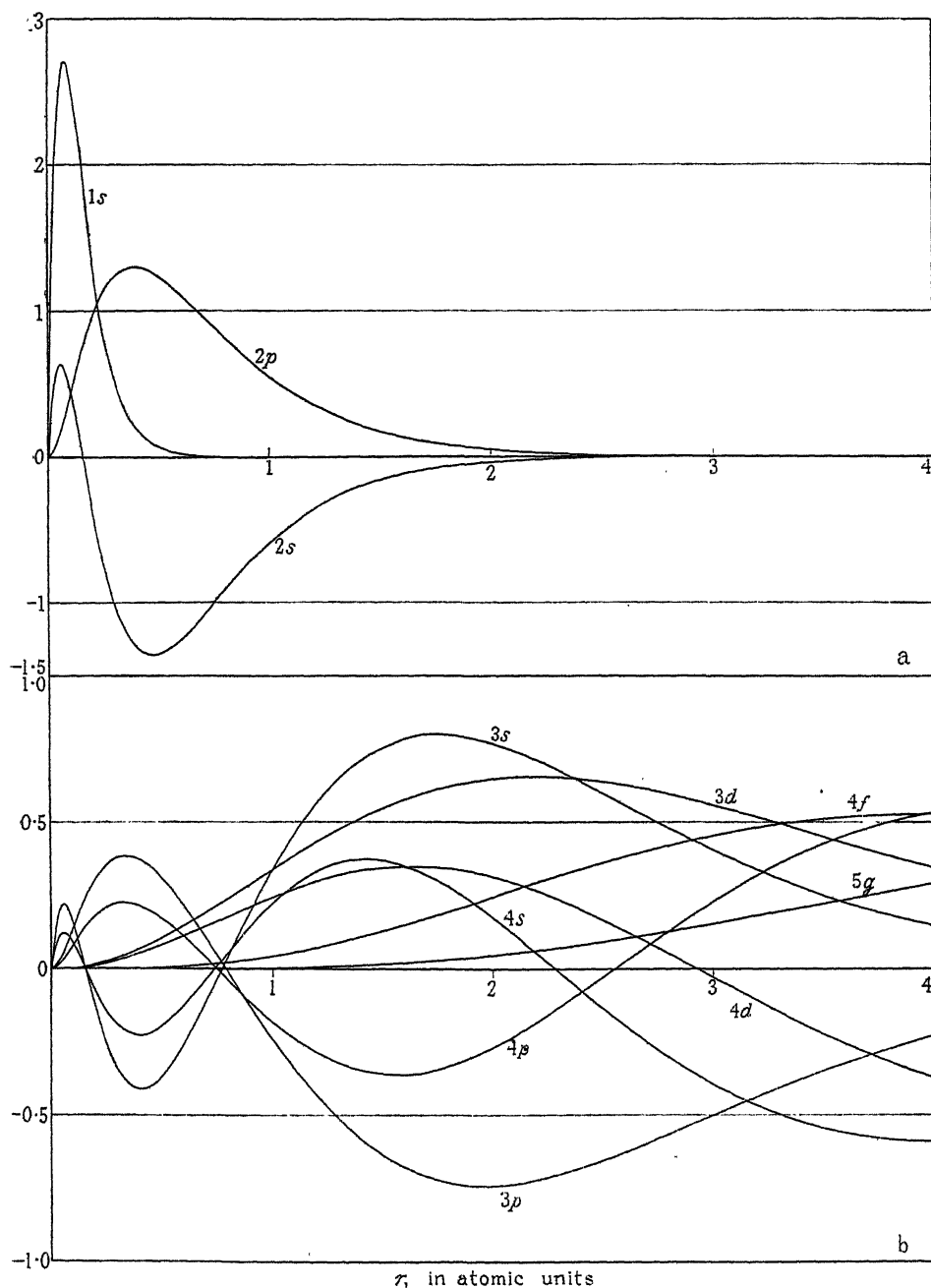
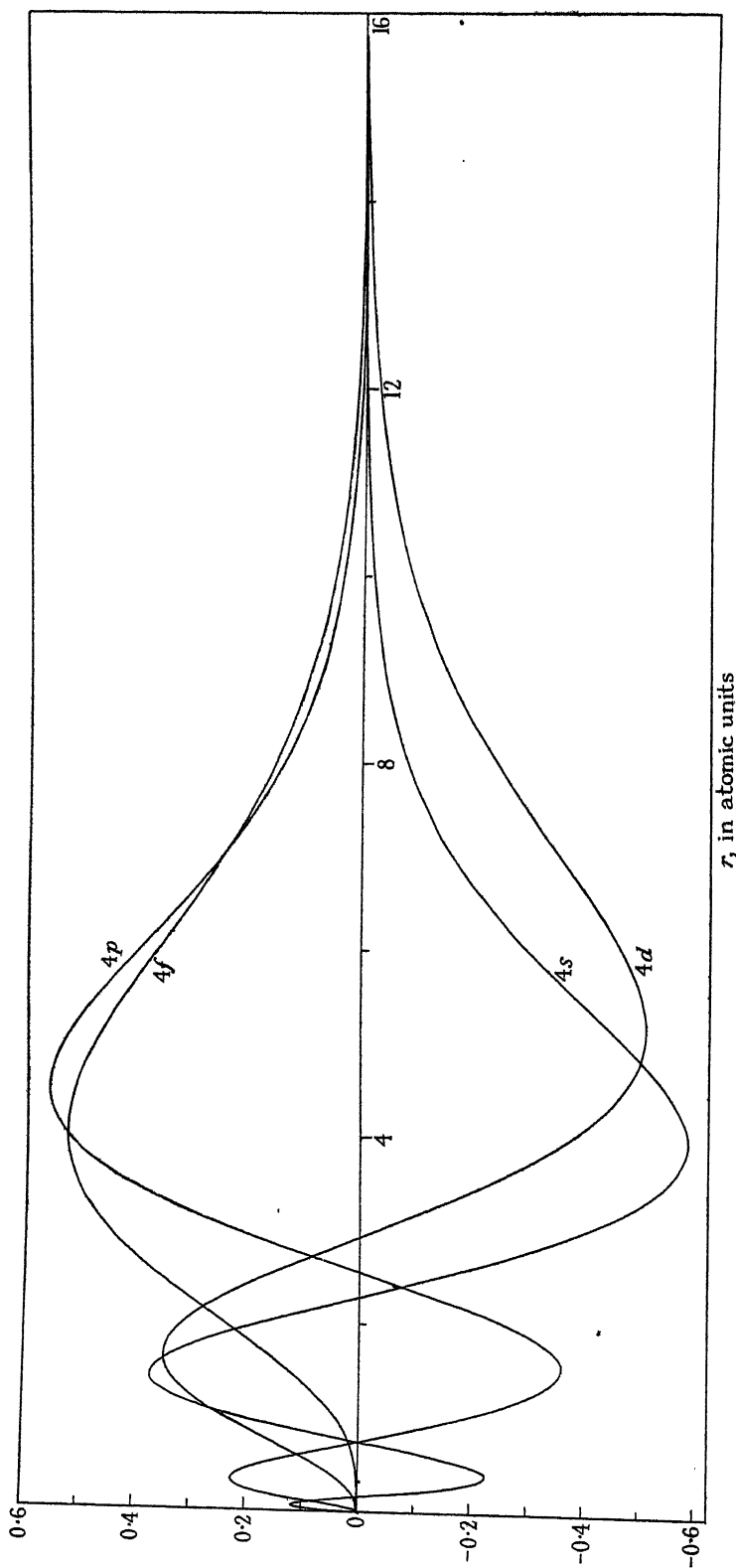


FIG. 1, *a* and *b*.—The normalised radial wave functions $P(\alpha|r)$ of the various central field problems. The range of nuclear distance r is the effective range of integration for the integrals calculated in this paper. Fig. 1, *a* shows the core wave functions, and the series electron wave functions are given in fig. 1, *b*. The scale of the ordinates in 1, *a* is half that in 1, *b*.



r , in atomic units

FIG. 1, c.—The normalised radial wave functions $P(\alpha|r)$ with quantum number $n = 4$. The curves are drawn for the whole range of r in which the functions differ appreciably from zero. (The principal maxima of $|P|$ for the wave functions with $n = 5$ occur in the neighbourhood of $r = 7, 8, 9$ and 6 for the $5s, 5p, 5d$ and $5g$ functions, respectively.)

tion of the integrals calculated in this paper. In fig. 1, c is shown the extent of the wave functions with quantum number $n = 4$.

A comparison of the observed and calculated ϵ values for 13 optical terms of Si^{+3} is made in Table II.* The calculated values of ϵ differ from the observed

Table II.—Optical Terms of Si^{+3} .

$$R = 109737. \quad n - q = 4/\sqrt{\epsilon}.$$

Term.	Observed ν .	Observed $\epsilon = \nu/R$.	Calc. ϵ .	ϵ obs. — ϵ calc.	q obs.	q calc.	q obs. — q calc.
$3s$	364092	3.318	3.096*	0.221*	0.804	0.726*	0.077*
$4s$	170114	1.5502	1.4847	0.0655	0.787*	0.717	0.070*
$5s$	98675	0.8992	0.8710	0.0282	0.782	0.714	0.068
$3p_2$	292340	2.6640	2.496	0.1680	0.549	0.468	0.081
$3p_1$	292804	2.6682		0.1722	0.551		0.083
$4p_2$	145664	1.3274	1.2728	0.0546	0.528	0.454	0.074
$4p_1$	145826	1.3288*		0.0560*	0.530		0.076
$5p_2$	87513.9	0.7975	0.7727*	0.0247*	0.521	0.450	0.071
$5p_1$	87589.4	0.7982		0.0254*	0.523		0.073
$6p_2$	58409	0.5322*	0.5189	0.0133*	0.517*	0.447	0.070*
$6p_1$	58452	0.5326*		0.0137*	0.519*		0.072*
Term.	Observed $\nu - \frac{C^2R}{n^2}$.	Observed ϵ .	Calc. ϵ .	ϵ obs. — ϵ calc.	q obs.	q calc.	q obs. — q calc.
$3d$	8631	1.8564*	1.8039	0.0525*	0.064 2*	0.021 8	0.042 4*
$4d$	4348	1.0396	1.0146	0.0250	0.076 9*	0.028 9	0.048 0*
$5d$	2371.3	0.6616	0.6483*	0.0132*	0.082 3*	0.032 3	0.050 0*
$4f$	230.0	1.0021	1.0002*	0.0018*	0.004 2	0.000 4*	0.003 7*
$5g$	25.5	0.6402*	0.6400	0.0002*	0.000 9	0	0.000 9
$6h$	5.0	0.4444 9	0.4444*	0.0000*	0.000 3	0	0.000 3

* I am indebted to Professor A. Fowler for allowing me to use some of his unpublished results of observations on the spectrum Si IV (second column of the table).

values by from 3 to 7 per cent. for the s and p terms, and up to 3 per cent. for the d terms. For the $4f$ term, the values agree within one-fifth per cent., and for the $5g$ and $6h$ terms the agreement is very close indeed.

The values of the quantum defect, $q = n - n^*$, *i.e.*, the difference between the principal and effective quantum numbers, are included in the table. (The quantum defect q and the energy parameter ϵ are related by the equation $C^2 = (n - q)^2 \epsilon$, where C is the core charge.) The principal reason for using this quantity is that it gives a measure of the departure of a term from hydrogen-likeness. The wave mechanics give results in agreement with experiment for a hydrogen-like atom, and the deviations of the observed terms of a many-electron atom from hydrogen-likeness are due to the interactions of the electrons of the atom, so that a proper test of the adequacy of any treatment of the many-electron problem is not a comparison of calculated and observed terms, but a comparison of the departures from hydrogen-likeness of the calculated and observed terms, and this departure is conveniently measured by the quantum defect. The difference between observed and calculated values of the *energy* may give an unduly favourable view of the accuracy of the approximation used for the many-electron problem. We observe from the table that the q values decrease as n increases for the s and p terms, but increase with n for the d , etc., terms, *i.e.*, the s and p terms are more hydrogen-like for increasing n , but the other terms diverge more from the hydrogen type as n increases.

The following points are worthy of note : (i) For any sequence of terms with the same l , the ϵ values vary greatly, while the q values are all of the same order of magnitude, *e.g.*, for the p terms taken, the ϵ values lie between 2.496 and 0.5189, while the q values are between 0.551 and 0.517'. (ii) The differences of energy for the members of the p doublets vary from 0.0042 to 0.0004, while the q differences are all 0.002 (actually 0.0020 or 0.0019'). (iii) The values of q obs. — q calc. for a sequence of terms are of the same order, *e.g.*, for the p terms, the range is 0.070 to 0.083. This assists when one is estimating the ϵ values. When the $3p$ calculation has been made, and its value of q obs. — q calc. obtained, an estimate of the $4p$ value of the q difference, and so of q calc. is made and the corresponding ϵ value is found. The number of approximations to find the required ϵ value is reduced in this way.

Hartree identified the calculated ϵ 's for the optical electrons with the optical terms, and attributed the discrepancies to a polarisation of the core by the series electron. Strictly, these ϵ numbers are only parameters in the central field problems of the series electrons, namely, the energies of one electron in an artificial static field, and are not the energies required to remove this electron

from the actual atom, whose core is not a static field and can only approximately be represented by one.

The accuracy of the wave functions of the series electrons is determined by the approximations which have been made. The first is the approximation involved in expressing the wave function of the atom in terms of the one-electron wave functions at all; both the forms (1.4) and (1.5) are only approximations to the true wave function of the atom. Further, the self-consistent field method effectively assumes the wave function for an atom to be of the form of a simple product such as (1.4), which is not nearly so good an approximation as the determinantal form (1.5). The use of the determinantal form leads to the Fock* equations, to which the Hartree equations may be considered to be first approximations, but the Hartree equations have been used rather than the more exact Fock equations on account of the technical difficulties involved in the solution of the latter. Further, the effect of the series electron in perturbing the core wave functions is neglected entirely. These approximations in the determination of core wave functions affect the series electron wave functions and ϵ values as calculated from the Hartree equations, and further changes would be introduced by the use of the Fock equations in the determination of the series electron wave functions and associated energy parameters.

We shall use some results of Slater,† who adopts the determinantal form of the atomic wave function, to express the energies of the whole atom in terms of the parameters ϵ of the central field problems and the corresponding one-electron wave functions.

3. Slater's Method for Atomic Energies.

The energy values are the diagonal terms of the matrix of the energy computed with respect to the wave function Ψ of the whole atom. If H is the energy operator, the diagonal term is

$$\int \Psi H \Psi d\tau, \quad (3.1)$$

assuming Ψ is normalised. Now the approximation to Ψ which we assume has the determinantal form

$$\Psi = \begin{vmatrix} \psi(\alpha|1) & . & . & . & \psi(\alpha|p) \\ . & . & . & . & . \\ \psi(\pi|1) & . & . & . & \psi(\pi|p) \end{vmatrix}, \quad (3.2)$$

* *Loc. cit.*

† *Loc. cit.*

where the individual functions $\psi(\alpha|j)^*$ are solutions of Schrödinger's equation for an electron in a central field, found by separation into spherical co-ordinates, viz.,

$$\psi(n_a, l_a, (m_l)_a | r_j, \theta_j, \phi_j) = R(n_a, l_a | r_j) S(l_a, (m_l)_a | \theta_j, \phi_j), \quad (3.3)$$

with the usual notation.[†] Slater follows Pauli in describing the spin, and sets up wave functions including space and spin co-ordinates alike, rather than a function involving these co-ordinates separately, thus

$$\begin{aligned} \psi(n_a, l_a, (m_l)_a, (m_s)_a | r_j, \theta_j, \phi_j, (m_s)_j) = & R(n_a, l_a | r_j) S(l_a, (m_l)_a | \theta_j, \phi_j) \\ & \times \delta((m_s)_a | (m_s)_j). \end{aligned} \quad (3.4)$$

The determinant Ψ is the sum of the expressions obtained by permutation of the co-ordinates of the electrons 1, 2, ... p in the product $\psi(\alpha|1) \dots \psi(\pi|p)$, and so the integral $\int \Psi H \Psi d\tau$ can be expressed in double sums. Slater assumes the one-electron wave functions are *orthogonal*, and so simplifies the integral, obtaining eventually the form

$$\sum_a I(n_a, l_a) + \sum_{\substack{\text{pairs} \\ a, \beta}} J(n_a, l_a, (m_l)_a; n_\beta, l_\beta, (m_l)_\beta) - \sum_{\substack{\text{pairs} \\ \parallel \text{ spins} \\ a, \beta}} K(n_a, l_a, (m_l)_a; n_\beta, l_\beta, (m_l)_\beta),$$

or shortly

$$\sum_a I(\alpha) + \sum_{a, \beta} J(\alpha; \beta) - \sum_{\substack{\parallel \text{ spins} \\ a, \beta}} K(\alpha; \beta), \quad (3.5)$$

the summations being taken over the wave functions (α) , (α, β) . The J and K integrals come from the interaction terms $1/r_{ik}$ in the energy operator, and the I integrals from the sums of terms involving the co-ordinates of one electron only. The integrals K only exist for pairs of electrons with parallel spins, and all the three types of integrals are expressed in terms of the central field wave functions. The $1/r_{ik}$ factors in J and K are expanded in surface harmonics; then these integrals may be written

$$J(n, l, m_l; n', l', m_{l'}) = \sum_k a_k(l, m_l; l', m_{l'}) F_k(n, l; n', l'), \quad (3.6)$$

$$K(n, l, m_l; n', l', m_{l'}) = \sum_k b_k(l, m_l; l', m_{l'}) G_k(n, l; n', l'), \quad (3.7)$$

* The Greek letter is the label of the wave function, and the numeral or Roman letter indicates the electron whose co-ordinates are substituted.

[†] The vertical stroke, e.g., in $R(n_a, l_a | r_j)$ is used to separate the symbols which are part of the label of the wave function from those which indicate the current co-ordinates of which R is a function. The semicolon is used to separate two sets of labelling numbers, e.g., in $J(n_a, l_a, (m_l)_a; n_\beta, l_\beta, (m_l)_\beta)$, and later in $S(\alpha; \beta)$.

where the coefficients a_k, b_k are certain definite integrals involving Legendre functions (see Table IV). By putting $rR(n_a, l_a|r) = P(n_a, l_a|r) = P(\alpha|r)$,* where R is the radial part of the wave function $\psi(\alpha|r)$, the integrals F_k, G_k may be written

$$F_k(\alpha; \beta) = F_k(n_a, l_a; n_\beta, l_\beta) = \int_0^\infty \int_0^\infty P^2(\alpha|r) P^2(\beta|r_1) \frac{r_a^k}{r_b^{k+1}} dr dr_1, \quad (3.8)$$

$$G_k(\alpha; \beta) = G_k(n_a, l_a; n_\beta, l_\beta) \\ = \int_0^\infty \int_0^\infty P(\alpha|r) P(\beta|r) P(\alpha|r_1) P(\beta|r_1) \frac{r_a^k}{r_b^{k+1}} dr dr_1, \quad (3.9)$$

where r_a is the smaller and r_b the greater of r and r_1 . The corresponding expression for I is

$$I(\alpha) = \int_0^\infty P(\alpha|r) \left\{ -\frac{1}{2} \frac{d^2}{dr^2} - \frac{N}{r} - \frac{1}{2} \frac{l(l+1)}{r^2} \right\} P(\alpha|r) dr. \quad (3.10)$$

Thus the energy for the whole atom is

$$E = \sum I(\alpha) + \sum_{\text{pairs}} J(\alpha; \beta) - \sum_{\substack{\text{pairs} \\ ||^{\text{spins}}}} K(\alpha; \beta), \quad (3.11)$$

where I, J, K are integrals involving the one-electron wave functions.

4. The Calculation of Energies of Optical Terms using Slater's Results.

The calculation of energy values may be divided into (i) the calculation of the approximate one-electron wave functions, and (ii) the calculation of energy values for the whole atom from these wave functions. The integral $\int \Psi H \Psi d\tau$

for the energy is stationary for small variations of the wave function used from the true wave function, and so we hope to get a good approximation to the energy from a fair approximation to the wave function. We obtain fair approximations to the one-electron wave functions, and so to the wave function for the atom, from the self-consistent field calculations. The one-electron wave functions are found by the methods of section 2, and are represented by such symbols as $P(3s|r)$.

The difference between the energies (as given by the equation (3.11)), for the configurations Si^{+3} and Si^{+4} , represents the work done in removing an electron from the Si^{+3} atom. This enables us to write the optical terms of Si^{+3} :

* Cf. equation (2.5). The symbol P is used to distinguish the *orthogonal* wave functions from the non-orthogonal wave functions P of § 2.

Energy of optical term λ ,

$$E_{\lambda} = I(\lambda) + \sum_{\alpha} J(\alpha; \lambda) - \sum_{\substack{\alpha \\ ||^{\text{spin}}}} K(\alpha; \lambda), \quad (4.1)$$

where λ is the label of a series electron wave function, and the sums are taken over the core wave functions. This equation holds only if the perturbation of the core by the series electron is neglected, but this is an approximation we are making (see §2, *b*). The double sums in (3.11) reduce to single sums over the core wave functions, as all the terms in this equation not involving the wave function of the series electron cancel out. We shall now investigate the methods for evaluating the integrals which occur in this expression.

(a) *Orthogonality of the Wave Functions*.—Slater works with one-electron wave functions which are orthogonal, and the wave functions P of the self-consistent field do not satisfy this condition. Those of different l are orthogonal through the surface harmonic functions, and we make those of different n but the same l orthogonal by taking linear combinations. There is no need to orthogonalise the wave functions, but we can do so without altering the determinant Ψ , since taking linear combinations of rows or columns makes a determinant which is a multiple of the first, and we ensure the initial and final determinants are the same by normalising the new wave functions; but, by using the orthogonal wave functions, the evaluation of the integral $\int \Psi H \Psi d\tau$ is greatly simplified, and we are able to use Slater's results immediately. Thus the possibility of orthogonalising is a property of the wave function of an atom in its determinant form, and we take advantage of this to simplify the work, not of necessity.

The normalised non-orthogonal wave functions are written P , and the normal and orthogonal functions, \mathbf{P} . For the s terms, we choose

$$\mathbf{P}(1s|r) = P(1s|r),$$

and we set

$$\mathbf{P}(2s|r) = A_1 P(2s|r) - A_2 \mathbf{P}(1s|r),$$

and choose A_1, A_2 so as to make $\mathbf{P}(2s|r)$ orthogonal to $\mathbf{P}(1s|r)$, and normalised. This requires

$$\mathbf{P}(2s|r) = \{P(2s|r) - AP(1s|r)\}/(1 - A^2)^{\frac{1}{2}},$$

where $A = \int_0^{\infty} P(2s|r) P(1s|r) \cdot dr$. The wave functions $\mathbf{P}(ns|r)$, where

$n > 2$, must be orthogonal to both $P(2s|r)$ and $P(1s|r)$. This is satisfied by

$$P(ns|r) = \{P(ns|r) - AP(2s|r) - BP(1s|r)\} / (1 - A^2 - B^2)^{\frac{1}{2}},$$

where

$$A = \int_0^\infty P(ns|r) P(2s|r) dr, \quad \text{and} \quad B = \int_0^\infty P(ns|r) P(1s|r) dr.$$

For the p terms, we set $P(2p|r) = P(2p|r)$, and make $P(np|r)$, for $n > 2$, orthogonal to $P(2p|r)$. The d, f, g wave functions are already orthogonal to the core wave functions through the surface harmonics. Thus, we may write the new wave functions in the form

$$\left. \begin{aligned} P(ns|r) &= \alpha P(ns|r) - \beta P(2s|r) - \gamma P(1s|r) \\ P(np|r) &= \alpha P(np|r) - \beta P(2p|r) \\ P(nd|r) &= P(nd|r), \text{ etc.} \end{aligned} \right\} \quad (4.2)$$

The orthogonalising integrals $S(\alpha;\beta) = \int_0^\infty P(\alpha|r) P(\beta|r) dr$, are given in Table III, and the coefficients β, γ are found in Table VII. Some of the normalised and orthogonal radial wave functions $P(\alpha|r)$ are graphed in fig. 2, where the core wave functions are drawn in the same figure as the series electron wave functions so that comparisons may be made more easily. The broken

Table III.— Si^{+3} Orthogonalising Integrals.

ns .	$S(ns;1s)$.	$S(ns;2s)$.	np .	$S(np;2p)$.
$2s$	0.0288		$3p$	0.0664*
$3s$	0.0047*	0.0396	$4p$	0.0310
$4s$	0.0032*	0.0172*	$5p$	0.0190*
$5s$	0.0017*	0.0102*	$6p$	0.0135

curves show the core wave functions, $1s$ and $2p$ being the same as in fig. 1, a ; but the $2s$ orthogonal curve is different from the non-orthogonal graph. This is easily seen in comparing the figures as the height of the maximum in the orthogonal curve is less than that in the corresponding P curve. The portion of the curve with positive ordinate is flatter than the corresponding portion of the fig. 1, a curve, and the zero has shifted slightly. Beyond $r = 0.5$, the non-orthogonal and orthogonal curves are identical. The changes in the series electron wave functions due to orthogonalisation are quite marked, *e.g.*, in the $3p$ curve, the maximum falls to about three-quarters of the value in the P curve, and the zero shifts appreciably. The curves differ until a point just beyond the minimum is reached (about $r = 2.3$), when they become identical. Similar differences will be noticed on comparing other corresponding curves in

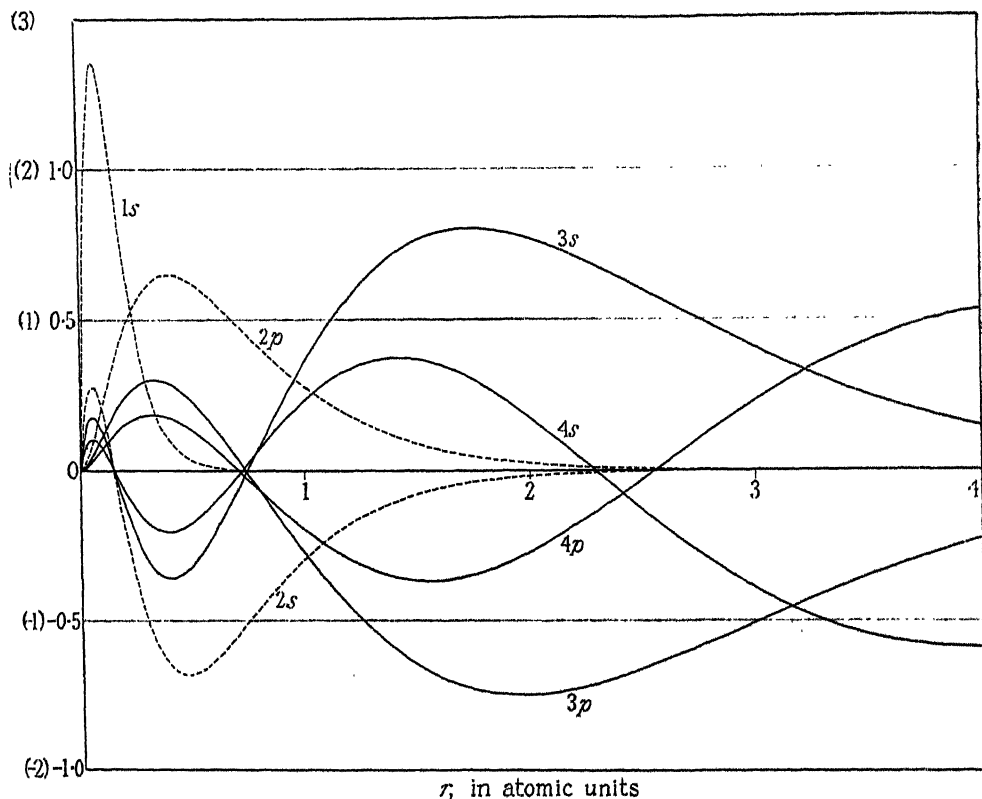


FIG. 2.—The orthogonal and normalised radial wave functions $P(\alpha|r)$. The broken curves show the core wave functions, and the series electron wave functions are indicated by full curves. The scale of the ordinates of the broken curves is indicated in parentheses. A comparison of figs. 1, *a* and *b* and fig. 2 shows the effect of orthogonalising. The *d*, *f*, *g* curves are not affected by orthogonalising, and are not repeated in fig. 2.

figs. 1, *b* and 2. The *d*, *f*, *g* wave functions are not drawn in fig. 2 as they are identical with those in fig. 1, *b*.

The *I*, *J*, *K* integrals are expressed in terms of the *P*'s, but some simplification is obtained by using the equation (2.6) for the *P*'s from which the orthogonal *P* functions are constructed.

(*b*) *The Interaction Integrals*.—The interaction integrals *J*, *K* have been written in terms of *F* and *G* integrals :

$$J(\alpha;\beta) = \sum_k a_k F_k(\alpha;\beta), \quad (4.3)$$

$$K(\alpha;\beta) = \sum_k b_k G_k(\alpha;\beta). \quad (4.4)$$

For the optical terms, we require $\sum_{\alpha} J(\alpha;\lambda)$, *i.e.*, the sum of the right-hand side of (4.3) over the wave functions of the core (and similarly for *K*).

The coefficients a_k , b_k depend upon the quantum numbers l , m_l of the series electron and core wave functions, and Slater (*loc. cit.*) has tabulated some of them. Those relating to interactions between f , g electrons and s , p electrons have been calculated for this work, and are given in Table IV. For the work on Si^{+3} , the summation is over the core configuration $(1s)^2(2s)^2(2p)^6$, and the total interactions of a series electron with these three shells have been collected in Table V. The functions there have been abbreviated as it is clear from the table to which wave functions they refer. The terms from J reduce to the F_0 's only since the series electron is outside complete shells.

Table IV.— $a_k(l, m_l; l', m_{l'})$ and $b_k(l, m_l; l', m_{l'})$ values for sf , sg , pf and pg interactions.

$$\begin{aligned} & (i) \quad a_k(l, m_l; l', m_{l'}) \\ a_k(l, m_l; l', m_{l'}) = & \frac{(2l+1)(l-|m_l|)!}{(l+|m_l|)!} \frac{(2l'+1)(l'-|m_{l'}|)!}{(l'+|m_{l'}|)!} \\ & \times \int_0^\pi [P_l^{|m_l|}(\cos \theta)]^2 [P_k^0(\cos \theta)] \frac{\sin \theta}{2} \cdot d\theta \\ & \times \int_0^\pi [P_{l'}^{|m_{l'}|}(\cos \theta')]^2 [P_k^0(\cos \theta')] \frac{\sin \theta'}{2} \cdot d\theta'. \end{aligned}$$

For sf and sg interactions—

$$(a) \quad a_0(0, 0; l', m_{l'}) = 1.$$

$$(b) \quad a_k(0, 0; l', m_{l'}) = 0, \text{ for all } k \neq 0.$$

For pf and pg interactions—

$$(a) \quad a_0(l, m_l; l', m_{l'}) = 1.$$

$$(b) \quad a_k(l, m_l; l', m_{l'}) = 0, \text{ except when } k = 0 \text{ or } 2.$$

Values of $a_2(l, m_l; l', m_{l'})$ are given by the table below.

Where there are two \pm signs, the two may be combined in any of the four possible ways.

<i>p.</i>		<i>pf.</i>			<i>pg.</i>		
<i>l.</i>	<i>m_l.</i>	<i>l'.</i>	<i>m_{l'}.</i>	<i>a₂.</i>	<i>l'.</i>	<i>m_{l'}.</i>	<i>a₂.</i>
1	± 1				4	± 4	4/55
1	± 1	3	± 3	1/15	4	± 3	1/55
1	± 1	3	± 2	0	4	± 2	— 8/385
1	± 1	3	± 1	—1/25	4	± 1	—17/385
1	± 1	3	0	—4/75	4	0	—4/77
1	0				4	± 4	— 8/55
1	0	3	± 3	—2/15	4	± 3	— 2/55
1	0	3	± 2	0	4	± 2	16/385
1	0	3	± 1	2/25	4	± 1	34/385
1	0	3	0	8/75	4	0	8/77

Any sequence a_k with m_l fixed and $m_{l'}$ changing by 1, has constant second order differences.

$$(ii) \quad b_k(l, m_l; l', m_{l'}).$$

$$b_k(l, m_l; l', m_{l'}) = \frac{(k - |m_l - m_{l'}|)!}{(k + |m_l - m_{l'}|)!} \frac{(2l + 1)(l - |m_l|)!}{(l + |m_l|)!} \frac{(2l' + 1)(l' - |m_{l'}|)!}{(l' + |m_{l'}|)!} \\ \times \left[\int_0^\pi P_l^{|m_l|}(\cos \theta) P_{l'}^{|m_{l'}|}(\cos \theta) P_k^{|m_l - m_{l'}|}(\cos \theta) \cdot \frac{\sin \theta}{2} \cdot d\theta \right]^2.$$

For *sf* and *sg* interactions—

$$b_k(0, 0; l', m_{l'}) = \frac{1}{2l' + 1}, \text{ when } k = l', \\ = 0, \text{ otherwise.}$$

For *pf* interactions—

$$b_k(l, m_l; l', m_{l'}) = 0, \text{ except when } k = 2 \text{ or } 4.$$

For *pg* interactions—

$$b_k(l, m_l; l', m_{l'}) = 0, \text{ except when } k = 3 \text{ or } 5.$$

Where there are two \pm signs, the two upper or the two lower signs must be taken together.

<i>p.</i>		<i>pf.</i>				<i>pg.</i>			
<i>l.</i>	<i>m_l</i>	<i>l'.</i>	<i>m_{l'}</i>	<i>b₂</i>	<i>b₄</i>	<i>l'.</i>	<i>m_{l'}</i>	<i>b₃</i>	<i>b₅</i>
1	± 1					4	± 4	4/21	1/363
1	± 1	3	± 3	9/35	1/189	4	± 3	1/7	1/121
1	± 1	3	± 2	6/35	1/63	4	± 2	5/49	2/121
1	± 1	3	± 1	18/175	2/63	4	± 1	10/147	10/363
1	± 1	3	0	9/175	10/189	4	0	2/49	5/121
1	± 1	3	∓ 1	3/175	5/63	4	∓ 1	1/49	7/121
1	± 1	3	∓ 2	0	1/9	4	∓ 2	1/147	28/363
1	± 1	3	∓ 3	0	4/27	4	∓ 3	0	12/121
1	± 1					4	∓ 4	0	15/121
1	0					4	± 4	0	3/121
1	0	3	± 3	0	1/27	4	± 3	1/21	10/363
1	0	3	± 2	3/35	4/63	4	± 2	4/49	7/121
1	0	3	± 1	24/175	5/63	4	± 1	5/49	8/121
1	0	3	0	27/175	16/189	4	0	16/147	25/363

Σb_k for given *l*, *l'*, *m_l* is independent of *m_{l'}*.

Any sequence *b_k* with *m_l* fixed and *m_{l'}* changing by 1, has constant second order differences.

Table V.—Interactions of series electron with the configuration $(1s)^2(2s)^2(2p)^6$ in terms of the F, G integrals.

Series electron $n > 2$.	Interaction with—		
	1s shell.	2s shell.	2p shell.
ns	$2F_0 - G_0$	$2F_0 - G_0$	$6F_0 - G_0$
np	$2F_0 - \frac{1}{3}G_1$	$2F_0 - \frac{1}{3}G_1$	$6F_0 - G_0 - \frac{2}{3}G_1$
nd	$2F_0 - \frac{1}{6}G_2$	$2F_0 - \frac{1}{6}G_2$	$6F_0 - \frac{2}{3}G_1 - \frac{6}{35}G_2$
nf	$2F_0 - \frac{1}{7}G_3$	$2F_0 - \frac{1}{7}G_3$	$6F_0 - \frac{2}{35}G_2 - \frac{2}{35}G_3$
ng	$2F_0 - \frac{1}{9}G_4$	$2F_0 - \frac{1}{9}G_4$	$6F_0 - \frac{2}{35}G_3 - \frac{5}{35}G_4$

(c) *Evaluation of the F, G Integrals.*—The calculation of these integrals is reduced to the solution of linear differential equations and quadratures.* We have

$$F_k(\alpha; \beta) = \int_0^\infty \int_0^\infty [P(\alpha|r)]^2 [P(\beta|r_1)]^2 \frac{r_a^k}{r_b^{k+1}} dr dr_1, \quad (4.5)$$

$$G_k(\alpha; \beta) = \int_0^\infty \int_0^\infty P(\alpha|r) P(\beta|r) P(\alpha|r_1) P(\beta|r_1) \frac{r_a^k}{r_b^{k+1}} dr dr_1, \quad (4.6)$$

α and β are labels of wave functions; r_a and r_b are respectively the smaller and greater of r, r_1 . Put

$$\rho(\alpha, \beta|r) = P(\alpha|r) P(\beta|r), \quad (4.7)$$

$$Y_k(\alpha, \beta|r) = r^{-k} \int_{r_1=0}^r \rho(\alpha, \beta|r_1) r_1^k dr_1 + r^{k+1} \int_{r_1=r}^\infty \rho(\alpha, \beta|r_1) r_1^{-(k+1)} dr_1, \quad (4.8)$$

and

$$Z_k(\alpha, \beta|r) = r^{-k} \int_{r_1=0}^r \rho(\alpha, \beta|r_1) r_1^k dr_1. \quad (4.9)$$

Then

$$\begin{aligned} \frac{Y_k(\alpha, \beta|r)}{r} &= \int_{r_1=0}^r \rho(\alpha, \beta|r_1) \frac{r_1^k}{r^{k+1}} dr_1 + \int_{r_1=r}^\infty \rho(\alpha, \beta|r_1) \frac{r^k}{r_1^{k+1}} dr_1 \\ &= \int_0^\infty \rho(\alpha, \beta|r_1) \frac{r_a^k}{r_b^{k+1}} dr_1. \end{aligned}$$

Hence

$$F_k(\alpha; \beta) = \int_{r=0}^\infty Y_k(\alpha, \alpha|r) \frac{\rho(\beta, \beta|r)}{r} dr, \quad (4.10)$$

and

$$G_k(\alpha; \beta) = \int_{r=0}^\infty Y_k(\alpha, \beta|r) \frac{\rho(\alpha, \beta|r)}{r} dr, \quad (4.11)$$

* I am indebted to Professor Hartree for this method of evaluating the double integrals.

and so we can evaluate these integrals when ρ and Y_k are known. A table of ρ values is prepared from tables of P , but it is not convenient to evaluate Y_k directly, particularly for the larger values of k , as the powers of r^{-1} and r occurring in the Y_k integral make the calculation very tedious for small r . In the evaluation of the first integral in (4.8), the number of decimal places to be retained changes so rapidly on account of the power of r^{-1} outside the integral and of r in the integrand. The powers of r^{-1} in the second integral make the integrand vary very quickly, and small intervals of integration would be necessary in order to evaluate the integral at all accurately. The integral is then to be multiplied by a power of r , which will have a small value. The calculation of Y_k by this method is clearly very troublesome. The method adopted is to obtain a differential equation between Y_k and Z_k :

$$\frac{dY_k}{dr} = \frac{1}{r} \{(k+1) Y_k - (2k+1) Z_k\}. \quad (4.12)$$

The procedure is: Evaluate the table of $Z_k(\alpha, \beta | r)$ values using equation (4.9), by integration outwards, and then calculate the $Y_k(\alpha, \beta | r)$ table by numerical integration of the differential equation (4.12) inwards. F_k and G_k are then obtained by quadrature. The inwards integration of (4.12) is started where r is so large that ρ is negligible; then the initial conditions are

$$\left. \begin{aligned} Y_k &= r^{-k} \int_0^\infty \rho(\alpha, \beta | r_1) r_1^k dr_1 = Z_k \\ \frac{dY_k}{dr} &= \rho - k \frac{Y_k}{r} \end{aligned} \right\}. \quad (4.13)$$

The powers of r inside the integral (4.9) for Z_k , and of r^{-1} outside, make the calculation of Z_k rather troublesome for small r as the number of decimal places to be retained changes so rapidly (especially for the larger values of k). We overcome this difficulty by finding a differential equation:

$$\frac{dZ_k}{dr} = \rho - k \frac{Z_k}{r}, \quad (4.14)$$

obtained by differentiating equation (4.9). Outwards integration of this equation is a practicable method and much more convenient than the direct evaluation of the Z_k integral for small r . For the larger values of k , it is convenient to use this method of finding Z_k for the whole range of integration.

This procedure has been adopted in the evaluation of the G integrals, but

it was not necessary to work out the F 's as only F_0 's appear, and they cancel with part of the I integrals.

(d) *The I and F_0 Integrals.*—The calculation of the I integrals is simplified by expressing the P functions which occur in them in terms of the P 's, and substituting the expressions for d^2P/dr^2 from the equation (2.6), which is satisfied by the P 's. The treatment of these integrals is shown in the following example :

$$I(3s) = \int P(3s|r) \left\{ -\frac{1}{2} \frac{d^2}{dr^2} - \frac{N}{r} + \frac{1}{2} \frac{l(l+1)}{r^2} \right\} P(3s|r) dr, \quad (4.15)$$

where

$$P(3s|r) = \alpha P(3s|r) - \beta P(2s|r) - \gamma P(1s|r), \quad (4.2)$$

and N is the atomic number, which is 14 here. Now $P(2s|r)$ is a linear combination of $P(2s|r)$ and $P(1s|r)$, [*i.e.*, $P(1s|r)$], and may be written

$$P(2s|r) = \alpha' P(2s|r) - \gamma' P(1s|r).$$

Thus

$$\begin{aligned} P(3s|r) &= \alpha P(3s|r) - \beta \alpha' P(2s|r) - (\gamma - \beta \gamma') P(1s|r), \\ &= AP(3s|r) - BP(2s|r) - CP(1s|r), \text{ say.} \end{aligned} \quad (4.16)$$

The wave function $P(\alpha|r)$ satisfies the equation

$$\left\{ -\frac{1}{2} \frac{d^2}{dr^2} + \frac{1}{2} \frac{l(l+1)}{r^2} \right\} P(\alpha|r) = \{v_\alpha - \frac{1}{2}\epsilon_\alpha\} P(\alpha|r), \quad (4.17)$$

where v_α is the potential of the field for which the function $P(\alpha|r)$ was calculated. [*Cf.* equations (1.3) and (2.6).] On substituting the expression (4.16) in equation (4.15), and using the relation (4.17) we obtain

$$\begin{aligned} I(3s) &= -\frac{1}{2} \int P(3s|r) \left\{ \left(\frac{2N}{r} - 2v_{3s} + \epsilon_{3s} \right) AP(3s|r) \right. \\ &\quad \left. - \left(\frac{2N}{r} - 2v_{2s} + \epsilon_{2s} \right) BP(2s|r) - \left(\frac{2N}{r} - 2v_{1s} + \epsilon_{1s} \right) CP(1s|r) \right\} dr. \end{aligned}$$

By making the $AP(3s|r)$ factor of the first round bracket up to $P(3s|r)$, this equation becomes

$$\begin{aligned} I(3s) &= -\frac{1}{2} \int \left(\frac{2N}{r} - 2v_{3s} + \epsilon_{3s} \right) P^2(3s|r) dr \\ &\quad - \int (v_{2s} - v_{3s}) P(3s|r) BP(2s|r) dr - \int (v_{1s} - v_{3s}) P(3s|r) CP(1s|r) dr \\ &\quad + \frac{1}{2} \int (\epsilon_{2s} - \epsilon_{3s}) P(3s|r) BP(2s|r) dr + \frac{1}{2} \int (\epsilon_{1s} - \epsilon_{3s}) P(3s|r) CP(1s|r) dr. \end{aligned}$$

Table VI.—Si⁴⁺. G Integrals and Contributions from them to Energies of Optical Terms.

Series electron.	G integrals.				Contributions to energy.				Sum of contributions.
	$G_0 (ns; 1s).$	$G_0 (ns; 2s).$	$G_1 (ns; 2p).$	$G_2 (ns; 2p).$	$G_0 (ns; 1s).$	$G_0 (ns; 2s).$	$G_1 (ns; 2p).$	$G_2 (ns; 2p).$	
<i>ns</i>									
3s	0.0139	0.0321*			0.0139	0.0321*			0.087
4s	0.0044 6*	0.0096 4*			0.0044 6*	0.0096 4*			0.0267*
5s	0.0023 7*	0.0043 3*			0.0023 7*	0.0043 3*			0.0124*
<i>np</i>									
	$G_1 (np; 1s).$	$G_1 (np; 2s).$	$G_0 (np; 2p).$	$G_2 (np; 2p).$	$\frac{1}{3} G_1 (np; 1s).$	$\frac{1}{3} G_1 (np; 2s).$	$\frac{1}{3} G_0 (np; 2p).$	$\frac{2}{3} G_2 (np; 2p).$	
3p	0.0194 9	0.0261 9	0.0319 5	0.0314 5	0.0064 9*	0.0087 3	0.0319 5	0.0125 8	0.0597*
4p	0.0077 1	0.0094 7	0.0113 4	0.0115 3	0.0025 7	0.0031 5*	0.0113 4	0.0046 1	0.0217
5p	0.0037 4*	0.0044 5	0.0052 0	0.0054 8	0.0012 5	0.0014 8*	0.0052 0	0.0021 9	0.0101
6p	0.0020 8	0.0024 5	0.0028 7*	0.0030 1*	0.0006 9	0.0008 1	0.0028 7*	0.0012 0	0.0056
<i>nd</i>									
	$G_2 (nd; 1s).$	$G_2 (nd; 2s).$	$G_1 (nd; 2p).$	$G_3 (nd; 2p).$	$\frac{1}{3} G_2 (nd; 1s).$	$\frac{1}{3} G_2 (nd; 2s).$	$\frac{2}{3} G_1 (nd; 2p).$	$\frac{5}{3} G_3 (nd; 2p).$	
3d	0.0000 6*	0.0214 3	0.0278 8	0.0156 6*	0.0000 1	0.0042 8*	0.0111 5	0.0040 3	0.0195
4d	0.0000 3*	0.0111 5	0.0141 4	0.0080 3	0.0000 0*	0.0022 3	0.0056 5*	0.0020 6*	0.0099
5d	0.0000 2	0.0061 4	0.0076 6*	0.0042 2*	0.0000 0*	0.0012 3	0.0030 6*	0.0010 8*	0.0054
<i>nf</i>									
	$G_3 (nf; 1s).$	$G_3 (nf; 2s).$	$G_2 (nf; 2p).$	$G_4 (nf; 2p).$	$\frac{1}{3} G_3 (nf; 1s).$	$\frac{1}{3} G_3 (nf; 2s).$	$\frac{2}{3} G_2 (nf; 2p).$	$\frac{5}{3} G_4 (nf; 2p).$	
4f	0	0.0002 8	0.0004 0*	0.0002 4	0	0.0000 4	0.0001 0	0.0000 4*	0.0001 8*
<i>ng</i>									
	$G_4 (ng; 1s).$	$G_4 (ng; 2s).$	$G_3 (ng; 2p).$	$G_5 (ng; 2p).$	$\frac{1}{3} G_4 (ng; 1s).$	$\frac{1}{3} G_4 (ng; 2s).$	$\frac{2}{3} G_3 (ng; 2p).$	$\frac{5}{3} G_5 (ng; 2p).$	
5g	0	0.0000 02	0.0000 04	0.0000 03	0	0.0000 00 2	0.0000 00 7*	0.0000 00 4*	0.0000 01*

Now $B = \beta\alpha'$, and by taking out the factor β and making the $\alpha'P(2s|r)$ factors up to $P(2s|r)$ and using the orthogonality properties, we obtain

$$I(3s) = -\frac{1}{2}\epsilon_{3s} - \int \left(\frac{N}{r} - v_{3s} \right) P^2(3s|r) dr - \beta \int (v_{2s} - v_{3s}) P(3s|r) P(2s|r) dr \\ - \gamma \int (v_{1s} - v_{3s}) P(3s|r) P(1s|r) dr - \beta\gamma' \int (v_{2s} - v_{1s}) P(3s|r) P(1s|r) dr.$$

Writing $(Z_p)_{\alpha}$, with round brackets, for the "effective nuclear charge for potential" for the field v_{α} , used in calculating the wave function $P(\alpha|r)$, we have

$$I(3s) = -\frac{1}{2}\epsilon_{3s} - \int \frac{N - (Z_p)_{3s}}{r} P^2(3s|r) dr - \beta \int \frac{(Z_p)_{2s} - (Z_p)_{3s}}{r} \rho(3s, 2s|r) dr \\ - \gamma \int \frac{(Z_p)_{1s} - (Z_p)_{3s}}{r} \rho(3s, 1s|r) dr \\ - \beta\gamma' \int \frac{(Z_p)_{2s} - (Z_p)_{1s}}{r} \rho(3s, 1s|r) dr, \quad (4.18)$$

The first integral in (4.18) cancels with part of the F_0 integrals; the contributions from these integrals will be called (F_0 & I) integrals. The remaining integrals with coefficients β , γ , etc., are called the (Remainder of I) Integrals. For a p wave function, γ and γ' are zero, and

$$I(3p) = -\frac{1}{2}\epsilon_{3p} - \int \frac{N - (Z_p)_{3p}}{r} P^2(3p|r) dr - \beta \int \frac{(Z_p)_{2p} - (Z_p)_{3p}}{r} \rho(3p, 2p|r) dr. \quad (4.19)$$

For the d, f, g functions, β , γ , and γ' are all zero.

(d) (i) (*Remainder of I*) Integrals.—The difference $(Z_p)_{2s} - (Z_p)_{3s}$ is the difference of the values of Z_p actually used in the self-consistent field equations for $P(2s|r)$ and $P(3s|r)$. Using $[Z_p]_{n,l}$, with square brackets, to indicate the contributions to Z_p from one n, l electron, as in § 2 (a), and similarly $[Z_p]_{\text{ion}}$, to indicate the contribution from the whole ion Si^{+4} , we have for $P(3s|r)$,

$$(Z_p)_{3s} = [Z_p]_{\text{ion}}.$$

For $P(2s|r)$, we have

$$(Z_p)_{2s} = [Z_p]_{\text{ion}} + 1 - [Z_p]_{2s};$$

thus

$$(Z_p)_{2s} - (Z_p)_{3s} = 1 - [Z_p]_{2s}. \quad (4.20)$$

The (Remainder of I) Integrals, such as

$$\beta \int \frac{(Z_p)_{2s} - (Z_p)_{3s}}{r} \rho(3s, 2s|r) dr,$$

are now found by quadrature, using equation (4.20) and the ρ values which were tabulated in the evaluation of the G integrals, and the values obtained for the s and p terms are given in Table VII. These integrals do not occur for the nearly-hydrogen like terms as the wave functions have not been changed by orthogonalisation, *i.e.*, the coefficients β , γ , etc., are zero.

Table VII.— Si^{+3} . (Remainder of I) Integrals.

ns .	$(ns;1s)$.			$(ns;2s)$.			$\beta\gamma'$ integral.			Sum of contributions.
	In- tegral	γ .	Contri- bution.	In- tegral.	β .	Contri- bution.	In- tegral.	$\beta\gamma'$.	Contri- bution.	
3s	0.266	0.0047*	0.0012*	0.159*	0.0396*	0.0063	-0.251	0.0011*	-0.0003	0.0072*
4s	0.154	0.0032*	0.0005	0.084*	0.0172*	0.0014*	-0.146	0.0005	-0.0001	0.0018*
5s	0.105*	0.0017*	0.0002	0.055	0.0102*	0.0005*	-0.097	0.0003	0	0.0007*

np .	$(np;2p)$.		Contribution.
	Integral.	β .	
3p	0.175	0.0666	0.0116*
4p	0.100	0.0310	0.0031
5p	0.068	0.0190*	0.0013
6p	0.049*	0.0135	0.0006*

(d) (ii) (F_0 & I) Integrals.—The F_0 integrals are of the form

$$\begin{aligned}
 F_0(\alpha; \beta) &= \int_0^\infty \frac{P^2(\alpha|r)}{r} \left\{ \int_{r_1=0}^r P^2(\beta|r_1) dr_1 + \int_{r_1=r}^\infty \frac{r}{r_1} P^2(\beta|r_1) dr_1 \right\} dr, \\
 &= \int_0^\infty \frac{P^2(\alpha|r)}{r} Y_0(\beta\beta|r) dr, \\
 &= \int_0^\infty \frac{P^2(\beta|r)}{r} Y_0(\alpha\alpha|r) dr, \text{ on integration by parts.} \quad (4.21)
 \end{aligned}$$

For a series electron wave function λ , the contribution from the F_0 integrals to the energy (see Table IV) is

$$2F_0(\lambda;1s) + 2F_0(\lambda;2s) + 6F_0(\lambda;2p)$$

$$= \int_0^\infty \{2Y_0(1s,1s|r) + 2Y_0(2s,2s|r) + 6Y_0(2p,2p|r)\} \frac{P^2(\lambda|r)}{r} dr. \quad (4.22)$$

The part of the I integral to be calculated with the F_0 's is

$$\int_0^\infty \frac{(Z_p)_\lambda - N}{r} P^2(\lambda|r) dr;$$

but

$$(Z_p)_\lambda = [Z_p]_{\text{ion}} = 2 [Z_p]_{1s} + 2 [Z_p]_{2s} + 6 [Z_p]_{2p} + 4,$$

for a series electron λ of Si^{+3} . (The 4 is the net positive charge on the Si^+ ion.) Thus the contribution from I is

$$\int_0^\infty \{2 [Z_p]_{1s} + 2 [Z_p]_{2s} + 6 [Z_p]_{2p} - 10\} \frac{\mathbf{P}^2(\lambda|r)}{r} dr. \quad (4.23)$$

Now when an exact solution of the self-consistent field problem has been obtained,

$$1 - [Z_p]_\alpha = \int_{r_1=0}^r P^2(\alpha|r_1) dr_1 + \int_{r_1=r}^\infty \frac{r}{r_1} P^2(\alpha|r_1) dr_1, \quad (4.24)$$

so that $\{1 - [Z_p]_\alpha\}$ bears the same relation to $P(\alpha|r_1)$ as $Y_0(\alpha\alpha|r_1)$ does to $\mathbf{P}(\alpha|r_1)$. Hence

$$1 - [Z_p]_{1s} = Y_0(1s, 1s|r),$$

$$1 - [Z_p]_{2p} = Y_0(2p, 2p|r),$$

since $P(1s|r) = \mathbf{P}(1s|r)$, and $P(2p|r) = \mathbf{P}(2p|r)$. The expression (4.23) becomes

$$\int_0^\infty \{-2Y_0(1s, 1s|r) + 2[Z_p]_{2s} - 6Y_0(2p, 2p|r) - 2\} \frac{\mathbf{P}^2(\lambda|r)}{r} dr,$$

and the sum of the contributions from the F_0 and I integrals is

$$(F_0 + I)_\lambda = \int_0^\infty \{2[Z_p]_{2s} - 2 + 2Y_0(2s, 2s|r)\} \frac{\mathbf{P}^2(\lambda|r)}{r} dr. \quad (4.25)$$

The $(F_0 + I)$ integrals may be calculated from this formula, or may be expressed entirely in terms of Y_0 functions or F_0 integrals:

From (4.24)

$$1 - [Z_p]_{2s} = \int_{r_1=0}^r P^2(2s|r_1) dr_1 + \int_{r_1=r}^\infty \frac{r}{r_1} P^2(2s|r_1) dr_1. \quad (4.26)$$

Now $P(2s|r)$ is a linear combination of $\mathbf{P}(2s|r)$ and $\mathbf{P}(1s|r)$, since § 4 (a) gives

$$\mathbf{P}(2s|r) = \alpha P(2s|r) - \beta \mathbf{P}(1s|r),$$

where

$$\alpha = 1/\sqrt{1-S^2}, \quad \beta = S/\sqrt{1-S^2}, \quad S = \int_0^\infty P(2s|r) \mathbf{P}(1s|r) dr.$$

Thus

$$\alpha^2 P^2(2s|r) = \mathbf{P}^2(2s|r) + \beta^2 \mathbf{P}^2(1s|r),$$

and equation (4.26) becomes

$$1 - [Z_p]_{2s} = \frac{1}{\alpha^2} Y_0(2s, 2s|r) + \frac{\beta^2}{\alpha^2} Y_0(1s, 1s|r).$$

By using the relation $\alpha^2 - \beta^2 = 1$, and the equation (4.21), the $(F_0 \text{ \& } I)$ integrals may now be written

$$\begin{aligned} (F_0 \text{ \& } I)_\lambda &= 2S^2 \int_0^\infty \{Y_0(2s, 2s|r) - Y_0(1s, 1s|r)\} \frac{P^2(\lambda|r)}{r} dr \\ &= 2S^2 \{F_0(2s; \lambda) - F_0(1s; \lambda)\}. \end{aligned} \quad (4.27)$$

The $(F_0 \text{ \& } I)$ integrals are small and negative, and arise from the fact that we orthogonalised the wave functions. The magnitudes of these integrals are given in Table VIII.

Table VIII.— Si^{+3} . $(F_0 \text{ \& } I)$ Integrals.

<i>ns</i> .	Integral.	<i>np</i> .	Integral.	Term.	Integral.
3s	0.0005*	3p	0.0003*	<i>nd</i>	0
4s	0.0002	4p	0.0001*	<i>nf</i>	0
5s	0.0001	5p	0.0001	<i>ng</i>	0
		6p	0.0000*		

In equation (4.1) we have expressed the energy E_λ of an optical term λ in the form

$$E_\lambda = I(\lambda) + \sum_a J(\alpha; \lambda) - \sum_a K(\alpha; \lambda), \quad (4.1)$$

the sum being over the core wave functions, supposed unperturbed by the presence of the series electron. For an atom with one electron outside complete shells, this has been reduced to

$$E_\lambda = I(\lambda) + \{\sum \alpha_0 F_0\}_\lambda - \{\sum b_k G_k\}_\lambda,$$

and in this section (4), the terms on the right-hand side have been written in a form involving the energy parameter ϵ_λ of the associated central field problem and certain integrals:

$$E_\lambda = -\frac{1}{2}\epsilon_\lambda - \{(\text{R. of I integrals})\}_\lambda - (F_0 \text{ \& } I)_\lambda \text{ integral} - \{\sum G \text{ integrals}\}_\lambda. \quad (4.28)$$

We have evaluated all the integrals occurring in equation (4.28) for E_λ , and are now able to apply the necessary corrections to the Hartree values $-\frac{1}{2}\epsilon_\lambda$ (Table II), and so find the real energies of the optical terms, to the approximations (3.2), (3.4) to the real wave function Ψ for the whole atom.

5. *The Real Energies of Optical Terms. Results and Discussion.*

In Table IX are collected the contributions to the real energies from the several integrals, for the 12 optical terms of Si^{+3} for which the calculations have been carried out. The "initial" E values of the table are those found in § 2 by the Hartree method (the calculated ϵ values are given in Table II, and $E = -\frac{1}{2}\epsilon$); and the "final" values of E are those found by applying the corrections to the initial E 's in accordance with equation (4.28). The improvement in the E values is very satisfactory, as the final values agree with the observed values to within 1 per cent. for terms with principal quantum number 3, $\frac{1}{2}$ per cent. for $n = 4$, and $\frac{1}{5}$ per cent. for $n = 5$.

The values of the quantum defect q are included in the table, as they provide the fairest test of the improvements effected by the methods of this paper (as already pointed out in § 2 (b)). The general improvement in the quantum defect values is very satisfactory, and there is an appreciable improvement in the variation of q in a sequence of terms with the same l , as well as in the absolute values. The values of q obs. — q calc. are reduced by factors of 7 or 8 or more for the s and p terms*; but the improvements for the nearly hydrogen-like terms are not so good, the q difference values for the d terms being reduced by a factor of 4 or 5 only, while the f and g terms are hardly affected. The reasons are to be found by recalling the corrections to the Hartree method which have not been taken into account.

In the determination of the self consistent field, the wave function for the atom was taken in the form of a simple product, and the effect of the series electron in perturbing the core—both the alteration of the core wave functions which appear in Ψ in $\int \Psi^* H \Psi d\tau$, and the perturbation of the series electron wave functions, which would be affected by the perturbation of the core, thus giving a polarisation effect on the series electron—was neglected entirely. Instead of using the Fock equations for the calculation of both core and series

* It may be pointed out that this shows that a better approximation to the wave function of the series electron outside the core is obtained by taking the calculated wave function in the self-consistent field of the core, than by taking the wave function with the observed ϵ (either in the coulomb field of the core joined up to the wave function inside, or in a field chosen to give the observed ϵ values, if possible); for the G integrals, etc., would be almost the same for both approximations, so that the properly calculated energy (which is the only test of the accuracy of the wave function adopted) would be much worse for the wave function with the observed ϵ than for the wave function calculated in the self-consistent field of the core.

Table IX.—Si⁺³. Real Energies of Optical Terms. Comparisons of Calculated and Observed Values.

Term.	Initial —E calc.	$\Sigma G's$.	Re. of I.	F ₀ & I.	Final —E calc.	—E obs.	Initial —E diff.	Final —E diff.	Initial q calc.	Final q calc.	q obs.	Initial q diff.	Final q diff.
3s	1.548	0.087	0.007	0.000	1.642	1.659	0.111	0.016	0.726	0.793	0.804	0.077	0.011
4s	0.7423	0.0267	0.0018	0.0002	0.7711	0.7751	0.0327	0.0039	0.717	0.779	0.787	0.070	0.008
5s	0.4355	0.0124	0.0007	0.0001	0.4438	0.4496	0.0141	0.0008	0.714	0.777	0.782	0.068	0.005
3p	1.2480	0.0597	0.0116	0.0003	1.3197	1.3327	0.0847	0.0129	0.468	0.538	0.550	0.082	0.012
4p	0.6304	0.0217	0.0031	0.0001	0.6613	0.6639	0.0275	0.0026	0.454	0.522	0.529	0.075	0.007
5p	0.3864	0.0101	0.0013	0.0001	0.3979	0.3988	0.0124	0.0009	0.450	0.516	0.521	0.071	0.005
6p	0.2594	0.0056	0.0006	0.0000	0.2657	0.2662	0.0067	0.0004	0.447	0.513	0.518	0.071	0.004
3d	0.9019	0.0195	—	0	0.9214	0.9282	0.0262	0.0067	0.022	0.053	0.064	0.042	0.010
4d	0.5073	0.0099	—	0	0.5172	0.5198	0.0125	0.0025	0.029	0.067	0.077	0.048	0.010
5d	0.3242	0.0054	—	0	0.3296	0.3308	0.0066	0.0012	0.032	0.073	0.082	0.050	0.009
4f	0.5001	0.0001	—	0	0.5003	0.5010	0.0009	0.0007	0.000	0.001	0.004	0.003	0.003
5f	0.3200	0	—	0	0.3200	0.3201	0.0001	0.0001	0	0	0.000	0.000	0.000

electron wave functions, Hartree's equations, which are first approximations to the Fock equations, have been used throughout. The result is that the wave functions (figs. 1 and 2) are slightly displaced (not uniformly) from the more exact Fock functions. The maxima and minima and zeros of the Fock wave functions will occur at values of r differing slightly from those shown in the drawings and this will change the r values at which the ρ functions are stationary or change sign and lead to modifications in the values of the integrals found in § 4.* For the s and p wave functions, the changes in the G integrals, say, are not likely to be large, as the P and ρ functions have comparatively large numerical values throughout most of the effective range of integration (up to $r = 4$),† and a slight shift in the ρ curves will not greatly change the comparatively large G integrals. For the nearly hydrogen-like terms, one would expect the modifications due to using the more exact wave functions would be greater proportionally, and more so for the f and g than for the d functions. The values of the G functions for the f and g terms are very small because the f and g wave functions do not differ appreciably from zero until such large values of r have been reached that the core wave functions are small (figs. 1 and 2), and so the ρ , Y and Z functions are very small throughout the range of integration. The effect of taking the more exact approximations indicated above would be to modify the core wave functions and the wave functions for the series electrons, and would probably increase the ρ functions and G integrals, due to the modified wave functions having larger values in the regions where they overlap.

To the final $-E$ calc. values of Table IX, corrections should be made for spin and relativity, for which formulæ in atomic units have been given by Hartree.‡ These calculations have not been made for this paper, but for the doublet p terms, the $-E$ obs. values have been corrected for spin, and the centre of gravity values have been inserted in Table IX. (The doublet values are given in Table II.) The spin energy§ is $A\{j(j+1) - l(l+1) - s(s+1)\}$, and has the values $-2A$ and A for $l = 1, j = \frac{1}{2}$ and $l = 1, j = \frac{3}{2}$ respectively.

* The changes will not be large as the integral $\int \Psi H \Psi d\tau$ is stationary for small departures from the correct wave function, so presumably small changes will be obtained for small departures from an approximate wave function.

† This is the range for which the radial wave functions are drawn in figs. 1, a and b , and fig. 2.

‡ 'Proc. Camb. Phil. Soc.,' vol. 24, p. 89 (1928).

§ Or $A[J(J+1) - L(L+1) - S(S+1)]$ for an atom with several electrons in incomplete groups, Goudsmit, 'Phys. Rev.,' vol. 31, p. 946 (1928).

Now the term of smaller j lies deeper*, *i.e.*, has the greater negative energy or greater term value, so that the term corrected for spin lies nearer the lesser term, the value being lesser term $+\frac{1}{2}$ difference. Thus, instead of correcting the calculated $-E$ value, we find the centre of gravity of the observed doublet for comparison with the calculated term value. To the $-E$ calc. values, a relativity correction should be made; this is of the same order of magnitude as the spin separations.

The calculations for the energies of optical terms given in this paper give results agreeing very well with the observed values, the discrepancies all being within 1 per cent. To account for the small residual differences, we have suggested certain inadequacies in our approximations, including a polarisation effect which we have neglected. It appears that the polarisation of the core by the series electron was not the true explanation of the discrepancies between the observed optical terms and the calculated ϵ values, such as was suggested by Hartree, as we have accounted for 6/7 of this difference for the $3s$ term, *e.g.*, without taking any account of a polarisation effect.

The idea of the polarisability of atom cores was used by Born and Heisenberg† and Swirles‡ to explain the values of optical terms on the orbital quantum theory, and on this model, Hartree§ showed that an additional field with an inverse fourth power potential αr^{-4} , will give energy values for any one sequence of nearly hydrogen-like terms of lithium-like and sodium-like atoms in close agreement with the observed values, but that the α value varied considerably with l . This is the type of field which would be produced by the polarisation of the core by the series electron. On the wave mechanics, Waller|| has calculated the nearly hydrogen-like terms of helium, by treating the interaction of the series electron and the core electron as a polarisation of the hydrogen-like He^+ core by the series electron.

The writer made some calculations of the self consistent field of Na^+ and of the optical terms of Na by the methods of § 2 and attempted to account for the discrepancies in the energy values by superposing on the self consistent field a "polarisation" field with an inverse fourth power potential at large distances. For small distances from the nucleus, this form of the potential was not expected as the field of the series electron is not uniform over the core, and the perturba-

* See Hund, 'Linienspektren,' pp. 74, 104.

† 'Z. Physik,' vol. 23, p. 388 (1924).

‡ 'Proc. Camb. Phil. Soc.,' vol. 23, p. 403 (1926).

§ 'Proc. Roy. Soc.,' A, vol. 106, p. 552 (1924).

|| 'Z. Physik,' vol. 38, p. 635 (1926).

tion of the core can probably not be represented by a doublet at the centre. It was necessary to use an empirical perturbing field of potential αr^{-4} , with α having a constant value for large r , but its behaviour for small r being chosen to give in combination with the self consistent field, the best fit for the observed terms. Some information as to the dependence of α on r was given by the nearly hydrogen-like terms, but it was not found possible to reconcile the observed and calculated values by the addition of such a field.

On the other hand, Guillemin and Zener* have made calculations on the normal state of lithium on wave mechanics, and have obtained good approximations without the explicit use of polarisation, but by using the proper form for the interaction between the series electron and the core. In this paper, we have obtained good approximations to the energies of the optical terms of an atom without the use of polarisation, and it appears that the magnitude of the polarisation effect must be very small. Of the points mentioned above which might account for the outstanding discrepancies, the use of the Fock equations is probably the most important, as it is a first order effect, whilst the polarisation effect is of the second order—and any further improvement in the results of this paper is likely to follow from the knowledge of a scheme for solving Fock's systems of equations by numerical methods.

I wish to thank Professor D. R. Hartree for his interest and encouragement throughout the course of this work, and the Department of Scientific and Industrial Research for a grant.

Summary.

Hartree's method of self consistent fields gives good energy values for the X-ray levels, but not for the optical terms, and this discrepancy has been attributed to a polarisation effect caused by the interaction of the series electron with the core. Some of Slater's results for atomic energies have been used to obtain corrections to the values found by the Hartree method. Numerical calculations for some optical terms of Si^{+3} have been made, and the agreement between observed and calculated energy values has been greatly improved. Inadequacies in the approximations made, which may account for the small residual differences, are discussed.

* 'Z. Physik,' vol. 61, p. 199 (1930).

Rotational Uncoupling, with application to the Singlet Hydrogen Bands.

By Dr. P. M. DAVIDSON, Senior Lecturer in Physics, University College, Swansea.

(Communicated by O. W. Richardson, F.R.S.—Received July 28, 1932.)

The general theory of uncoupling has been discussed by Kronig and Fujioka, and by MacDonald,* but before considering the hydrogen bands it will be convenient to establish some general results. We consider first the integral

$$\int P_{vKA} \cdot B \cdot P_{vKA'} \cdot \rho^2 d\rho, \quad (1)$$

which appears in the formulæ, and which may be called $B_{vvAA'}$. In the hydrogen levels we shall only consider the first one or two values of v and the first four or five values of K ; thus the unperturbed levels can be written as

$$W_A^0 = C_{(v)A} + B_{(v)A} (K + \tfrac{1}{2})^2 - D_{(v)A} (K + \tfrac{1}{2})^4 + F_{(v)A} (K + \tfrac{1}{2})^6 - \dots \quad (2)$$

The F term is exceedingly small at our K 's.

We consider first the value of (1) in the hypothetical case where the $U_n(\rho)$'s of the two states concerned differ only by a constant, so that the two P eigenfunctions are identical, and in (2) it is only the C 's which differ. The ordinary vibrational equation (MacDonald's (5) for example) may be regarded as containing $K(K+1)$ or $(K + \frac{1}{2})^2$ as a parameter; calling either of them ε we have the general formula

$$(W_A^0)_{\varepsilon+\Delta\varepsilon} = (W_A^0)_{\varepsilon} + (\partial W_A^0 / \partial \varepsilon)_{\varepsilon} \Delta\varepsilon + \text{term in } (\Delta\varepsilon)^2 + \dots \quad (3)$$

But the perturbation theory (for variation of a parameter) shows at once that in the case we are considering the coefficient of $\Delta\varepsilon$ in (3) is exactly the integral (1), whose value is thus found by differentiating (2). It is

$$B_{(v)} = 2D_{(v)} (K + \tfrac{1}{2})^2 + 3F_{(v)} (K + \tfrac{1}{2})^4 - \dots \quad (4)$$

Suppose now there are other differences between the two U_n 's; suppose, in

* Full references will be found in MacDonald's paper, 'Proc. Roy. Soc.,' A, vol. 138, p. 183 (1932). Symbols which I have not explained will be found in these papers, or in Kronig's book.

other words, that the constants B_0 , γ (see footnote *) and the a , b in the U series are somewhat different in the two states. Using any general type of U curve, and assuming that none of the above constants differ by more than about 10 per cent. in the two states, I find that at our v 's (say 0 and 1), the integral is†

$$\bar{B}_{(v)} - 2\bar{D}_{(v)}(K + \frac{1}{2})^2 - \frac{\bar{\nu}_0}{16} \left(\frac{\Delta B_0}{\bar{B}_0} \right)^2 (v + \frac{1}{2}) + \dots, \quad (5)$$

Here a bar denotes the average value for the two states. Substituting the approximate $\bar{B}_{(v)}$ and $\bar{\nu}_0$ of the states that we shall consider, it is found that if ΔB_0 (the difference of the B_0 's) is $\bar{B}_0/10$, the last term in (5) is about 8 per cent. of the whole at $v = 1$. At $v = 0$ it is less than 3 per cent. If the constants only differ by about 1 per cent. this last term may be omitted; in other words, the constants in (4) have simply to be replaced by their average values.

We consider next the behaviour of the perturbed levels. As before we assume at first that the unperturbed energies (2) differ only in their C's. The five perturbed levels at a given K are determined by the quadratic and cubic‡ in the paper by Kronig and Fujioka. Inserting the expression (4) above, we find that as K increases the roots tend at first towards

$$\bar{C} + B_{(v)}(K^{\times\times} + \frac{1}{2})^2 - D_{(v)}(K^{\times\times} + \frac{1}{2})^4 + \dots \quad (6)$$

Here \bar{C} has one value for Π_a , one for Σ and Δ_a , and one for Π_b and Δ_b ; but the important point is that $K^{\times\times}$ has the values $K - 2$, $K - 1$, K , $K + 1$, $K + 2$, in the five states Σ , Π_b , Π_a , Δ_b , Δ_a . (The physical meaning is well known.) The expression (6) is approached when the ω 's become fairly large compared with the C intervals. Using approximate values for the $4d^1$ states (which have smaller C intervals than the $3d^1$ states) it is found that at about $K = 6$ the roots agree with (6) to within about 20 wave-numbers. It is still better to replace $K^{\times\times}$ by K^\times , defined by

$$K^\times(K^\times + 1) = K(K + 1) + z, \quad (7)$$

* If desired, $B_{(v)}$ may be written as $B_0 + 3B_0\gamma(1 + a)(v + \frac{1}{2}) + \delta(v + \frac{1}{2})^2 + \dots$; γ is $2B_0/\nu_0$; a is the coefficient in $U_n(\rho) = \text{const.} + k\xi^2(1 + a\xi + b\xi^2 + \dots)$; $D_{(v)}$ is $B_0\gamma^2 + \delta'(v + \frac{1}{2}) + \dots$; B is $\hbar^2/8\pi^2\mu\rho^2$, and B_0 is the value at ρ_0 .

† MacDonald's expressions for the B integrals are designed to show their *orders of magnitude* for various values of v and v' , not, as above, to specify a fairly accurate value for the case of $v = v'$. They assume the Kratzer U_n , in which a has the value -2 .

‡ For the He_2 bands this would read "quadratic or cubic," since there are only two, or three, levels at a given K; (the rotational levels which Kronig and Fujioka mark with crosses, and which are weak in H_2 , are entirely absent in He_2).

where z is zero in Π_a , h_{12} in Δ_b , $-h_{12}$ in Π_b , $(h_{01}^2 + h_{12}^2)^{\frac{1}{2}}$ in Δ_a and minus the same* in Σ . Then for ω 's large compared with the C intervals the expression (6) differs from the roots of the quadratic and cubic by, firstly, an expression which, being due to the lack of uncoupling at small K , decreases as K increases and is zero if the three C 's in (2) are identical; secondly, by terms which increase with K , causing the roots to diverge from (6); and the largest term† of this type is $-z^2 D_{(v)}$, where z has the values above.‡ This last result, however, has no meaning, for we have here overstepped the limit of accuracy of Kronig and Fujioka's approximation: to improve it we must extend their simple determinants to take into account, in the first place, perturbations by the other vibrational levels. Considering, for example, Δ_b and Π_b , and taking first the hypothetical case where the U_n 's only differ infinitesimally, retaining their actual order $U_n(\Delta) > U_n(\Pi) > U_n(\Sigma)$,§ it will be found that the deter-

* h_{01} is $2[3K(K+1)]^{\frac{1}{2}}$ and h_{12} is $2[(K-1)(K+2)]^{\frac{1}{2}}$. Thus (7) can be written as $(K^{\times} + \frac{1}{2})^2 = (K + \frac{1}{2})^2 + z = (K^{\times \times} + \frac{1}{2})^2 + p + q/K^{\times \times} + \dots$, where p and q depend on the state, but not on $K^{\times \times}$, which has the integral values above. Π_b and Δ_b have a common p , and so have Σ and Δ_a .

† Provided K is not so large that the series (2) is ceasing to converge.

‡ Thus the term is approximately $-4D_{(v)}(K + \frac{1}{2})^2$ in Π_b and Δ_b ; and approximately $-16D_{(v)}(K + \frac{1}{2})^2$ in Σ and Δ_a .

§ I have defined them as *infinitesimally close* rather than *coincident* so as to retain an unambiguous correlation of the perturbed and unperturbed states. For $K \geq 2$, the equations of the theory merely determine five levels (each with its eigenfunctions and transition intensities) without connecting them in a definite way with the unperturbed states. Thus when the latter are closely adjacent as in the $4d^1$ levels (or still more in the present hypothetical case) the correlation is not immediately obvious. It might be made by neglecting the perturbations due to other vibrational levels, and imagining Kronig and Fujioka's ω 's reduced continuously to zero. Consider, for example, the cubic: its roots are given (fig. 1) by the intersection of the cubic $y = (W - W_0^0)(W - W_1^0)(W - W_2^0)$ and a straight line whose slope dy/dW is $(\omega_{01}^2 + \omega_{12}^2)$, and whose intersection with the W axis divides the line $W_0^0 W_1^0 W_2^0$ internally in the ratio $\omega_{01}^2 : \omega_{12}^2$. It is easily seen that for all real values of the ω 's the greatest root is always greater than any of the W^0 's, and eventually passes into the greatest of them as the ω 's decrease; the least root is always less than any of the W^0 's and passes into the least of them; and the third root always lies between these extreme W^0 's and passes into the central one. In these H_2 bands the order is $W_2^0 > W_1^0 > W_0^0$, giving the identifications in the text.

This diminution of the ω 's with constant W^0 's has no definite physical meaning (nor indeed have the W^0 's themselves); instead, then, we may imagine, for example, that the nuclear mass is continuously increased to ∞ . As before the slope of the line diminishes, but now also the W^0 's move at the same time, passing eventually into the minimum values of the W_b 's in equation (22) of Kronig's book. At every stage the greatest of the (fully perturbed) W 's is greater than any of the W^0 's, and passes eventually into the greatest of the W_b 's; and so forth. According to a paper by MacDonald, the order of our W_b 's

minant can be reduced to the product of two, which are exactly those which specify the energies of

$$\left[B \frac{\partial}{\partial \rho} \left(\rho^2 \frac{\partial}{\partial \rho} \right) - U_n(\rho) - BK^\times(K^\times + 1) + W \right] P_{\nu K^\times} = 0, \quad (8)$$

where $K^\times(K^\times + 1) = K(K + 1) \pm h_{12}$, exactly as in (6), and where the same equation with K^\times replaced by K specifies the unperturbed levels. The same result can be obtained very readily by a method due to MacDonald, which also gives readily the corresponding result for Σ , Π_a and Δ_a . All are contained in (8), K^\times having the values (7).^{*} Substituting the actual expressions for z we can expand $K(K + 1) + z$ in the form $K^{\times \times}(K^{\times \times} + 1) + p + q/K^{\times \times} + \dots$, where p and q are independent of $K^{\times \times}$, which has the integral values given previously. Thus as K increases, the level of, for example, Δ_a , differs from the unperturbed $K + 2$ level by a diminishing fraction of a rotational interval in that region.[†] (If the rotational intervals do not increase indefinitely, this difference will tend to zero.)

It is easily shown that the corresponding eigenfunctions[‡] are, for $K \geq 2$,

$$\left. \begin{aligned} \Psi(\Pi_a) &= \frac{P_{\nu K^\times}}{(h_{01}^2 + h_{12}^2)^{\frac{1}{2}}} [h_{12}F_0 - h_{01}F_2] \\ \Psi(\Pi_b) &= \frac{P_{\nu K^\times}}{2^{\frac{1}{2}}} [F_1 - F_2]; \\ \text{for } \Delta_b \text{ change the sign of } F_2. \\ \Psi(\Sigma) &= \frac{P_{\nu K^\times}}{2^{\frac{1}{2}}} \left[\frac{h_{01}F_0}{(h_{01}^2 + h_{12}^2)^{\frac{1}{2}}} - F_1 + \frac{h_{12}F_2}{(h_{01}^2 + h_{12}^2)^{\frac{1}{2}}} \right]; \\ \text{for } \Delta_a \text{ change the sign of } F_1. \end{aligned} \right\} \quad (9)$$

is the same as that of the W^0 's, so that in our case the two definitions give the same result. It is to be noted that in some of the lowest K 's the correlation is in any case unambiguous, so that unless the W^0 differences are decidedly small, changes in correlation at the higher K 's are likely to cause sharp discontinuities in some of the bands.

^{*} The modified (6), which has K^\times instead of $K^{\times \times}$ (and has here a single value for the three C 's, equal to the single value of the three C 's), represents these energies, until it ceases to converge. The terms spoken of above, of which the largest was $-z^2 D_{(v)}$, thus represent the perturbations by the other vibrational levels. Nor should these perturbations be altogether different in the actual levels, for the C differences are fairly small fractions of a vibrational interval, and theoretically the B_0 's should not differ much.

[†] Note that we have got away from the idea of a series for the W^0 's.

[‡] The same eigenfunctions (9) would be obtained from the expressions of Kronig and Fujioka (on setting the W^0 differences infinitesimal and the P 's identical), except that they would contain $P_{\nu K}$ instead of $P_{\nu K^\times}$. The small difference between these functions represents, of course, the perturbations by the other vibrational levels.

In each case the $P_{\sigma K^*}$ is the solution of (8) for that value of K^* which specifies the energy. Π_a is remarkable in that whereas its energy levels are entirely unperturbed, its eigenfunctions are perturbed to such an extent that they no longer contain their own unperturbed eigenfunctions. The latter result may sound paradoxical, but it must be remembered that the unperturbed states represent the limit when the W^0 differences are infinitely greater than the ω 's, while the Ψ 's above represent just the converse limit.* Nor are these Ψ 's merely of academic interest; they are nearly attained in the hydrogen states, especially, of course, in the larger K 's, and especially in the 4-electronic states (owing to the smallness of the W^0 differences). Thus even at $K = 3$ in $4d^1\Pi_a$, Kronig and Fujioka's S_{10} , S_{11} , S_{12} have values 0.47, 0.15, -0.87, so that the unperturbed eigenfunction plays quite a small part.

This hypothetical case of practically identical U_n 's may be regarded as a special example of a more general case, in which we take the three U_n 's to be of the form

$$U_n = f(\rho) + Bc_\Lambda, \quad (10)$$

where only the constant c_Λ depends on Λ . (This should be nearly true at fairly large values of the principal electronic quantum number n , where the W_b 's will not differ much.† c_Λ will then have a value about $-2\Lambda^2$, for according to MacDonald h'_Λ only contains Λ in a term $-B\Lambda^2$.) Assuming a U_n of the form (10) and taking account of the perturbations by all the vibrational levels, I find that the energies are the eigenwerte of

$$\left[B \frac{\partial}{\partial \rho} \left(\rho^2 \frac{\partial}{\partial \rho} \right) - f(\rho) - B\bar{K}(\bar{K} + 1) + W \right] \bar{P}_{\sigma\bar{K}} = 0. \quad (11)$$

Here $\bar{K}(\bar{K} + 1) = K(K + 1) + \beta$, where β is one of the five roots of a quadratic and a cubic determinantal equation, of the same form as Kronig and Fujioka's, but with β instead of W ; h_{01} and h_{12} instead of ω_{01} and ω_{12} ; c_0 , c_1 , c_2 instead of W_0^0 , W_1^0 , W_2^0 . (Remembering that Kronig and Fujioka's treatment is in all cases a first approximation, the meaning of the analogy is

* Near this limit the Schrodinger perturbation method (using a series beginning with the unperturbed function) is quite inapplicable, owing to the divergence of the series. It is only applicable near the other limit. A method given by MacDonald is always applicable.

† Using approximate electronic eigenfunctions by MacDonald, it can be shown that, to a first approximation, the W_b ' differences fall off as n^{-3} . (The same is true of levels given by the Rydberg formula $W'_{bn\Lambda} = A - B/(n + x_\Lambda)^2$, where x_Λ is a small fraction, independent of n .) The known states ($n = 3$ and 4) agree fairly well with this n^{-3} law. It is evident that for n greater than a certain value the order of the three W^0 's should become reversed.

easy to trace.) As K increases the levels differ from those obtained by giving the K in (II) the integral values $K^{\times\times}$ by a diminishing fraction of a rotational interval in that region.* For all K 's the eigenfunctions are $\Psi_A = (\sum_p S_{A,p} F_p) \bar{P}_{v\bar{K}}$, the S 's being given by Kronig and Fujioka's equations, on changing the symbols as before; p has the values 0, 1, 2 in the α states and 1, 2 in the b states. Apart from the difference in the rotational eigenfunctions, the present Ψ 's pass asymptotically into the previous (10) as K increases. In applying this method to the experimental levels, we should write the W of (11) in the form (2), with \bar{K} instead of K , and see whether it were possible to fit the data by assigning values to the differences of the c 's, and to, say, the first four constants in (2).

For a U_n of the form (10), and also, of course, for the special case of identical U_n 's, the error which would be caused by neglecting the perturbations due to the other vibrational levels is easily found by comparing the accurate W given by (11) with the W which Kronig and Fujioka's method would give in this particular case. As would be expected, we find again the term $-z^2 D_{(v)}$, mentioned above, followed by other terms which only amount to a fairly small fraction of it, provided the differences of the W^0 's (and therefore of the $B \cdot c_A$'s) are similarly small fractions of a vibrational interval, and provided the same is true of the whole perturbation of the W^0 , i.e., provided K is not too large.

As remarked above, the assumption (10) is justifiable at fairly large n 's, since the W_b ' differences are then small; it is also justifiable if they themselves are nearly of the form (10), especially in the region of ρ_0 . Actually this is probably not true; but the exact form is not known, and the effect on the results will not be discussed. Some approximate expressions (for which I am indebted to Dr. MacDonald) for the differences of the W_b 's suggest that in the states $\Lambda = 0, 1, 2$, the extreme ρ_0 's should not differ by more than about 3 per cent. If this is so, a failure of Kronig and Fujioka's method requiring a perturbation of a larger order than $-z^2 D_{(v)}$ can probably not, at our v 's, be ascribed to neglect of the perturbations by other vibrational levels.†

* Inserting the values $-2\Lambda^2$ for the c_A 's in the limiting case of $n = \infty$, we find, as we should expect, that every rotational level of the molecule coincides with a rotational level of the ion.

† In considering other perturbations, it is to be noted that some of the "unexplained" singlet levels lie close to those we are dealing with. For example, the $v = 0$ level of the "unexplained" 3^1K system lies about 150 w.n. below the $v = 0$ level of $3d^1\Sigma$, which incidentally it resembles considerably in structure (though lacking its tremendous strength); in fact it might almost be substituted for $3d^1\Sigma$. On the whole, however, there are strong reasons for retaining the present Σ level, at least at $v = 0$; at $v = 1$ and 2 there is something to be said for interchanging them.

Some other corrections, notably that due to the inaccuracy (discussed by MacDonald) of the expressions for h_{01} and h_{12} must be left outstanding. They can be treated separately.

We consider now the application to the singlet hydrogen bands. Most of them have already been published; the others, notably the Δ bands, which incidentally were discovered before the theory was applied to them, will be given in a separate communication by Professor O. W. Richardson and the author. Full references will be found there.

The essential correctness of the theory is illustrated by taking the unperturbed $v = 0$ levels to be

$$W_A^0 = C_A + 28(K + \frac{1}{2})^2 - \frac{1}{56}(K + \frac{1}{2})^4,$$

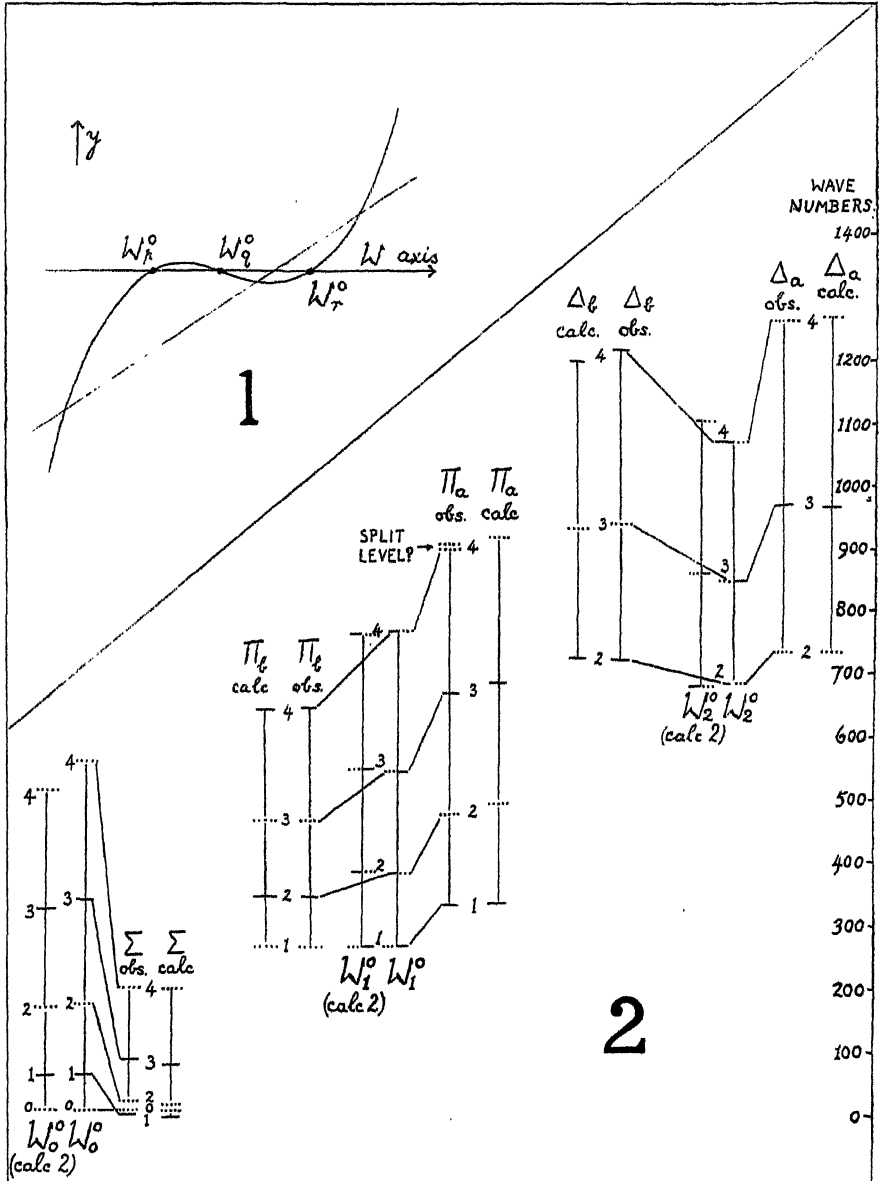
where C_A is $21601\frac{1}{2}$, $21804\frac{1}{2}$, $22111\frac{1}{2}$ for the 3-electronic states, and 27233, 27318, 27373 for the 4-electronic states. All these are measured upwards from the $K = 0$, $v = 0$ level of $2p^1\Sigma$. (In each of the diagrams an arbitrary zero is used.) The levels marked "calc." in figs. 2 and 3 are then found by means of the quadratic and cubic. Small corrections for the effects of the other vibrational levels have been made; even at $K = 4$ they are hardly perceptible on the scale of the diagrams. The agreement with the experimental levels is seen to be very fair,* and could probably be improved by a little adjustment; it is not likely, for example, that the $B_{(v)}$'s are all identical. We therefore adopt a different method of testing the theory; we find what small adjustments must be made in the values of the five constants W_0^0 , W_1^0 , W_2^0 , ω_{01} , ω_{12} , for a given K , in order that the five levels predicted by the quadratic and cubic shall agree with the experimental levels.† The number five applies at $K \geq 2$. At $K = 0$ and 1 there are, respectively, one and three perturbed levels, and the same number of adjustable constants.

This has been done for the 3-electronic states, since it is in them that the previous method shows the largest discrepancies. A small correction is made to allow for perturbation by the other vibrational levels; for this it is assumed that the ρ_0 's differ only slightly in the three states. The W_0 's so found are marked "calc. 2" in fig. 2. From each of the intervals between successive K levels of these W_0 's we can get, by using (2) and estimating the small

* It will be seen that some of the weak experimental levels have been omitted from fig. 2. Satisfactory sets of lines can be found, but being necessarily weak and fragmentary they are less convincing than the others.

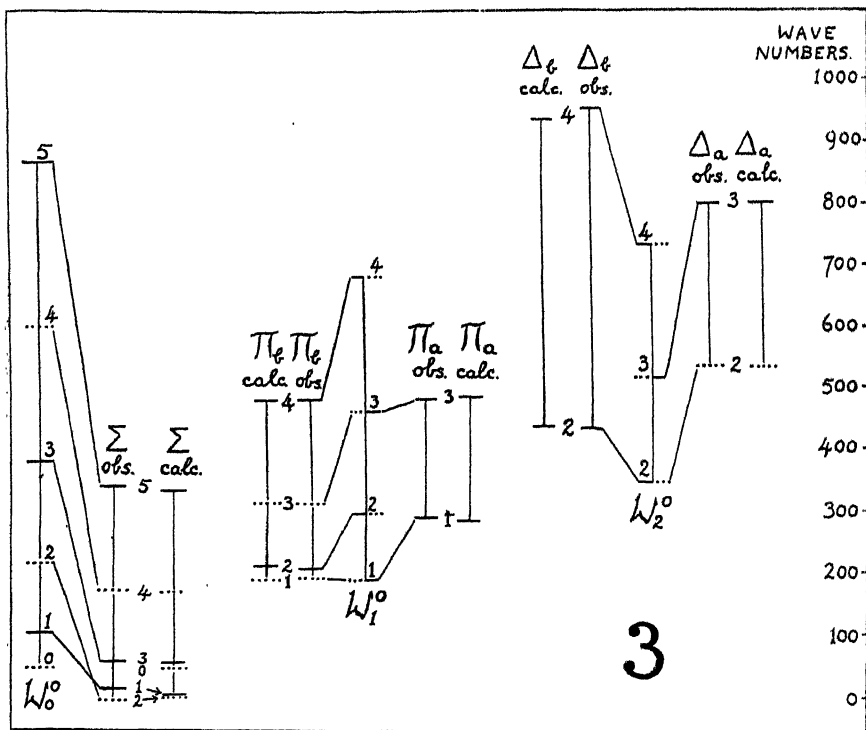
† This method cannot be used in the helium bands, owing to the absence of the alternate K levels of each band.

$(K + \frac{1}{2})^2$ term, a value of $B_{(v)\Lambda}$ for $v = 0, \Lambda = 0$. These values in order of increasing K are 27.9, 27.1, 25.8, 24.4. For the corresponding quantity at $\Lambda = 1$ we find 29.2, 28.1, 26.8; and at $\Lambda = 2$ we find 29.8, 31.1. If the



theory were accurate, the numbers in any one of these three sets would be identical, though, of course, there is no reason why the three sets should be identical. Each of the ω_{01} 's which accompany the W^0 's of "calc. 2" should

be equal to h_{01} times the quantity $B_{v\Lambda\Lambda'}$ discussed previously (v is 0; Λ and Λ' are 0 and 1). Estimating the value of the small $(K + \frac{1}{2})^2$ term in (5), we find from each of the ω_{01} 's a value for the part of (5) independent of K . The values, from the ω_{01} 's in order of increasing K , are 26.6, 26.5, 26.4, 26.0. Their lack of constancy indicates that the theory is inaccurate. Similarly from the ω_{12} 's we find for the constant part of B_{0012} the values 29.7, 28.9, 27.7. Essentially similar results are found if we omit the correction for the perturbation by the other vibrational levels. All things considered, it seems



likely that the ρ_0 's of $\Lambda = 0, 1, 2$ are in diminishing order of magnitude, and that there are perceptible perturbations not considered by the theory. The 4-electronic states are more satisfactory.

Tables I and II show the intensities. From a given K of $nd^1\Sigma$ there are, in general, five transitions—three to $2p^1\Pi$, and two to $2p^1\Sigma$. The relative intensities of these five lines can be compared theoretically. The same is true of $nd^1\Pi_a$ and $nd^1\Delta_a$. From $nd^1\Pi_b$, and also from $nd^1\Delta_b$, there are only four lines—three to $2p^1\Pi$ and one (Q) to $2p^1\Sigma$. In the helium bands considered by Kronig and Fujioka there were only transitions to the Π state; here we are comparing transitions to two different states.

The numbers under the headings x^2 are those given by the formulæ of Kronig and Fujioka, and of MacDonald, who is of the opinion that the relative intensities of the five (or four) lines should be given with considerable accuracy on multiplying each x^2 by the factor

$$v^4 \left[\int P_v' P_v'' \cdot \rho^2 \cdot d\rho \right]^2 \quad (12)$$

whose evaluation I have discussed in a different connection.* Its value, for a given upper level, will be nearly the same for the three lines to $2p^1\Pi$, but very different for the two lines (or one line) to $2p^1\Sigma$, the frequency being very much greater. In fact for $3d^1$ transitions the v^4 introduces a factor of about 8 in favour of the $2p^1\Sigma$ lines; for $4d^1$ lines the factor is about 5. The effect is somewhat counterbalanced by the integrals, for the potential energy curves of the upper states almost coincide with that of $2p^1\Pi$, whereas the curve of $2p^1\Sigma$ is considerably different (it has a very large ρ_0). The effect of the integrals is to introduce a factor of 2.0 in favour of the $2p^1\Pi$ lines. Thus altogether we have for $3d^1$ transitions a factor of about 4 in favour of $2p^1\Sigma$ lines, and for $4d^1$ transitions a factor of about $2\frac{1}{2}$. The numbers under the headings kx^2 are obtained by multiplying the x^2 's by the exact values of these factors, and at the same time multiplying the whole set of five (or four) lines by a factor which makes the intensity of the strongest line identical with the quantitative measurement of Kapuscinski and Eymers,† recorded in the next column under the heading E. In the remaining columns are the eye-estimates of Gale, Monk and Lee‡ (headed G), and of Poetker,§ (headed P.).

For the 4-electronic states (Table I) there are quantitative measurements throughout, and it will be seen that the agreement is satisfactory. The only definite failure is at $\Pi_a \rightarrow \Pi$, $K = 1$, Q, where, instead of the very weak line predicted by theory, a very strong line is found. This is presumably due to interference; there are a number of strong unassigned lines in this part of the spectrum. Such interference definitely causes the great strength of the line $\Sigma \rightarrow \Pi$, $K = 2$, R, which is required as a strong line in the Fulcher bands. There are also a few other cases of interference, mostly in fairly weak lines.

It is to be noted that the $2p^1\Pi$ state lies about 9000 wave-number above the $2p^1\Sigma$ state, so that the lines to these two states from a given upper level lie

* 'Proc. Roy. Soc.,' A, vol. 135, p. 459 (1932), and actual calculations by W. C. Price, 'Proc. Roy. Soc.,' A, vol. 136, p. 264 (1932).

† 'Proc. Roy. Soc.,' A, vol. 122, p. 58 (1929).

‡ 'J. Astrophys.,' vol. 67, p. 89 (1928).

§ 'Phys. Rev.,' vol. 30, p. 418 (1927).

Table I.

	K = 1.				K = 2.				K = 3.			
	x^2	kx^2	E.	G.	x^2	kx^2	E.	G.	x^2	kx^2	E.	G.
$4d^1\Sigma \rightarrow 2p^1\Sigma$ $\rightarrow 2p^1\Pi$	P R	0.6 3.1	17½ 109	(1) (7)	0.2 3.2	1 19	— 19	ab (1)	0.1 3.2	4 119	— 119	ab (5)
	P Q R	0 2.3	— 30	ab (5)	0 0.8 1.7	0 2 4	— 2.8 32.7†	ab (1) (10)†	0 0.4 2.3	0 5½ 35	— 4.9 <45.8	(0) (1) (9)
$4d^1\Pi_a \rightarrow 2p^1\Sigma$ $\rightarrow 2p^1\Pi$	P R	3.1 0.2	19.5 1.3	(2) (1)					0.8 0.15	14 2½	— ex	(0) ab
	P Q R	2.2 0.15	6 ½	(2) (6)*					0 3.0 1.9	0 22 14½	— 22 16.3	(4) (2)
$4d^1A_2 \rightarrow 2p^1\Sigma$ $\rightarrow 2p^1\Pi$	P R				2.6 0.1		ex —	(1a) ab	2.7 0	8 0	ex —	(1) ab
	P Q R				2.8 0.4 0.14		— — —	(1) ab ab	2.9 0.25 0	3.4 ½ 0	3.4 3.4† —	(0) (0)† ab

$4d^1\Pi_b \rightarrow 2p^1\Sigma$	Q	3	12.4	12.4	(0)	1.9	45.6	45.6	(3)	1.8	12.3	12.3	(1)
$\rightarrow 2p^1\Pi$	P	1.5	2 $\frac{1}{2}$	2.7	(0)	0.3	3	3.0	(0)	0.15	$\frac{1}{2}$	2.9	(1)
	Q	1.5	2 $\frac{1}{2}$	—	ab	0.1	1	3.5†	(00)†	0.35	1	2.0	(6)
	R					3.7	36 $\frac{1}{2}$	45.6	(6)	3.6	10	14.6†	(2)†
$4d^1\Delta_b \rightarrow 2p^1\Sigma$	Q					1.1	14 $\frac{1}{2}$	ex	(1)				
$\rightarrow 2p^1\Pi$	P					1.5	8	10.6†	(3)†				
	Q					2.4	13.4	13.4	(2)				
	R					0.85	4 $\frac{1}{2}$	4.7	(1)				

Two of the weak sets have been omitted, since they gave practically no evidence for or against the theory. "ex" means that the line lies outside the spectral range examined by the particular observer. In column G. (and P. in Table II) "ab" means that the line is too weak to be detected. In column E, — means that the line's intensity is not measured; it does not follow that it is very weak.

* See the text.

† Means that the line is also used in another system. The only important cases are $4d^1\Sigma \rightarrow 2p^1\Pi$, R, K = 2, and $4d^1\Pi_b \rightarrow 2p^1\Pi$, R, K = 3. The first is required in the Fulcher bands, where, however, it seems a little too strong. The second is also in the Fulcher bands. The strength required there is about half the strength of the line at 18497.14, which according to Gale is rather stronger than the present line; that, however, is probably a misprint, since the line is not given at all in the lists of Kapuscinski and Eymers, or of Merton and Barrat, both of which lists give the present line, most of the strength of which is therefore in the present system, probably.

‡ A little of this is required in another system.

Lines which coincide are indicated by a bar connecting them.

The original analysis of the $4d^1$ bands is here slightly re-arranged; the matter will be discussed in the paper dealing with the Δ bands referred to on p. 586.

Table II.

	K = 1.						K = 2.						K = 3.						K = 4.												
	E.			P.			E.			P.			E.			P.			E.			P.			G.			P.			
	x^2	kx^2		x^2	kx^2		x^2	kx^2		x^2	kx^2		x^2	kx^2		x^2	kx^2		x^2	kx^2		x^2	kx^2		x^2	kx^2		x^2	kx^2		
$3d^1\Sigma \rightarrow 2p^1\Sigma \dots$	P	1.1	87	(6)	ex	0.48	23	16.1	(3)	ex	0.25	36	25.1	(5)	ex	0.2	10	10	(2)	ex	0.55	6	ex	(1c) [†]	(3)	ex					
	R	2.8	227	(10)	ex	3.2	160	160	(8)	ex	3.2	485	485	(10)	ex	3.3	175	<146*	(5)	ex	1.9	23	ex	(0)	(1)	ex					
		0	0	ab	ab	0	0	ex	ab	ab	0	0	ex	ab	ab	ab	0.55	6	ex	ab	0.55	6	ex	(1c) [†]	(3)	ex					
$\rightarrow 2p^1\Pi \dots$	P	0	0	ab	ab	0	0	ex	ab	ab	0	0	ex	ab	ab	0.55	6	ex	ab	0.55	6	ex	(1c) [†]	(3)	ex						
	Q	2.1	40	(1)	(3)	1.2	14	ex	(06)	(1)	0.75	26	ex	(1)	(2)	1.9	23	ex	(2)	ex	1.9	23	ex	(0)	(1)	ex					
	R					1.1	13	ex	(0)	(1)	1.7	61	ex	(1)	(2)				(1)	ex											
$3d^1\Pi_a \rightarrow 2p^1\Sigma$	P	2.6	123	(10)	ex	2.6	73	73	(8)	ex	2.4	188	188	(10)	ex	2.1	31.4	31.4	(4)	ex	2.1	31.4	31.4	(4)	ex						
	R	0.56	27	(4)	ex	0.1	3	<11.4	(2h)	ex	0	0	0	ab	ex	0	0	—	ab	ex	0	0	—	ab	ex						
$\rightarrow 2p^1\Pi$	P	2.2	25	(006)	(2)	1.25	8	ex	(0)	(1)	0.6	11½	ex	(00)	(1h)	2.0	7½	ex	(1)	(1h)	2.0	7½	ex	(1)	(1)	ex					
	Q	0.45	5	(00)	ab	0.3	2	ex	ab	ab	1.3	25	ex	(1)	ex				(1)	ex											
	R					1.6	10	ex	(0)	(1)	1.65	33	ex	(1)	(2)	1.6	6	ex	ab	(2)	1.6	6	ex	ab	(1)	ex					
$3d^1A_a \rightarrow 2p^1\Sigma$	P					0.5	14.3	14.3	(2)	ex	0.95	34.2	34.2	(6)	ex					ex											
	R					0.1	3	3.2	(1)	ex	0.1	4	5.7	(0)	ex					ex											
$\rightarrow 2p^1\Pi$	P					1.4	10	ex	(0)	(1)	2.3	21	ex	(0)	U				U												
	Q					1.9	14	ex	(0)	ex	1.6	15	ex	(1)	(2)				(2)												
	R					2	15	ex	(1)	ex	0.8	8	ex	(1a)	ex				ex												
$3d^1\Pi_b \rightarrow 2p^1\Sigma$	Q	3	60	(5)	ex	2.7	286	286	(10)	ex	2.5	76	76	(6)	ex	2.4	13½	13½	(8)	ex	2.4	13½	13½	(8)	ex						
$\rightarrow 2p^1\Pi$	P	1.5	7	ab	ab	1.0	24	ex	(1)	(2)	0.7	5½	ex	ab	ab	0.1	1	ex	ab	0.1	1	ex	ab	(0) [†]							
	Q	1.5	7	ab	ab	0.1	3	ex	ab	ab	0.05	½	ex	ab	ab	1.8	25	ex	ab	1.8	25	ex	ab	(2)							
	R					2.0	52	ex	(1a)	(3)	2.6	20	ex	(1a)	ex				(1a)												
$3d^1A_b \rightarrow 2p^1\Sigma$	Q					0.3	33.7	33.7	(6)	ex	0.5	8.6	8.6	(2)	ex	0.6	26.4	26.4	(3)	ex	0.6	26.4	26.4	(3)	ex						
$\rightarrow 2p^1\Pi$	P	0.9	24	ex	(1)	0.9	24	ex	(1)	U, <(3)	1.5	6	ex	ab	ab	2.7	32	ex	ab	2.7	32	ex	ab	(2)							
	Q	2.5	71	ex	(3)	2.1	60	ex	(3)	(2)	2.7	11½	ex	ab	ab	0.8	10	ex	ab	0.8	10	ex	ab	(1a)	(2)	ex					
	R	2.1	60	ex	(2)			ex	(2)	ex	1.3	6	ex	ab	ab				ex												

Lines which Gale, Monk and Lee mark with the letter *a* were difficult to locate exactly; *b* and *c* even more so. *h* means that the line was definitely diffuse. The cause is presumably interference, so that the estimate is likely to be larger than the prediction. A question mark against one of Poetker's lines means that it has a combination defect exceeding his usual latitude of error; *U* against one of his lines means there is a combination defect probably due to lack of resolution from an adjacent line—again the estimate is likely to be greater than the prediction.

* E has not resolved this from an adjacent weaker line.

† Probably also in the Fulcher bands. In $3d^1H_b \rightarrow 2p^1\Pi$ the R lines, which are recorded by both G. and P., and give excellent combinations, are omitted from the original analysis by M. L. Chalk.

in widely different parts of the spectrum ; thus in using the eye-estimates of Gale, Monk and Lee it is only lines to the same state which should be compared.

In transitions from the 3-electronic states (Table II) the lines to $2p^1\Sigma$ show a good agreement with the quantitative measurements. The lines to $2p^1\Pi$ lie in the far infra-red, a region to which the quantitative measurements do not extend. The predicted relative intensities of the lines to this state agree fairly well with the eye-estimates, but in comparing them with the lines to $2p^1\Sigma$ we must make the quite probable assumption that Gale's eye-estimates in the far infra-red do not by any means correspond to those in the green, where the $2p^1\Sigma$ lines lie. Since the results for the 4-electronic states suggest that the theory is not likely to be far wrong, we may correlate Gale's infra-red eye-estimates with Kapuscinski's quantitative measurements. Considering all the lines to $2p^1\Pi$ in Table II, and taking the average† of the predicted intensities in the column headed kx^2 for each of the numbers which Gale uses, we find $ab = 3$, $(00) = 8$, $(0) = 14$, $(1) = 30$, $(2) = 60$, $(3) = 71$. Similarly for Poetker's numbers we find $ab = 2$, $(1) = 13$, $(2) = 36$, $(3) = 40$. We can then assign a range to each of Gale's numbers, and it is found that only about six lines do not fit into the scheme. The same is true of Poetker's ; moreover, most of the lines which are "wrong" according to one observer are "right" according to the other. In fact there are only two cases where both observers disagree with the theoretical intensity :— Δ_u , Q, K = 3, where the theoretical intensity is too small, and Σ , R, K = 3, where it is too large.

The tentative and uncertain nd^3 bands, outlined a year ago by O. W. Richardson and the author, do not fit the theory, indicating, no doubt, that some of them need revision.

I am indebted to Dr. J. K. L. MacDonald for some interesting correspondence, and to Mr. W. C. Price, of this College, for a graphical evaluation of the vibrational integrals of the type (12).

Summary.

Some points in the theory of rotational uncoupling are discussed, and it is applied to the nd^3 bands of hydrogen. The levels themselves and the intensities in the bands are found to agree fairly well with the theory.

† In computing both sets of averages I have discarded lines which Gale, who uses the higher resolving power, marks with letters a , b , c , h , indicating that they are indefinite, or abnormally broad. This is usually due to interference. (In taking the averages for Poetker's numbers I have omitted lines which he has not resolved from adjacent lines in Gale.) There remain 37 lines for computing the first list (Gale's) ; nevertheless, it will be noted that the values assigned to (2) and (3) each rest on a single line, and so have not much meaning. The same applies to the (3) in the Poetker list.

The Theory of Metals.—I.

By A. H. WILSON, Emmanuel College, Cambridge.

(Communicated by R. H. Fowler, F.R.S.—Received July 29, 1932.)

Introduction.

1. The quantum theory of electrical conduction in a solid has two main problems to face, the number of "free electrons" and the "mean free path." Of these the first is the simpler and has, to a certain extent, been solved. The evaluation of the mean free path, on the other hand, has given rise to some controversy and cannot be regarded as satisfactory. In his original paper on conduction Bloch† gave a theory of the interaction of the electrons and the thermal vibrations in a metal which leaves much to be desired from the point of view of rigour, but which leads to results in good agreement with experiment. Peierls‡ criticised this treatment and gave a new one, which, if correct, would considerably alter the theory.§ Peierls omitted most of the calculations, which are difficult, and based his treatment on physical arguments, which are by no means easy to follow, and which require justification. Recently L. Brillouin|| has given an extended mathematical treatment of the points in dispute, and obtains results which differ considerably from those of both Bloch and Peierls.

None of these calculations is really satisfactory, the main objection being that the physical assumptions have not been made sufficiently precise. A method is given here which treats consistently the interaction of the electrons and the lattice, and which enables the assumptions to be clearly seen. It also has the advantage that it can be extended quite naturally to deal with the problems of the dispersion and absorption of light in metals, which will be treated in subsequent papers. In this paper the general theory will be developed, and applied to the discussion of the debatable points in the theories of Bloch and Peierls. Although the general opinion seems to be that Peierls' criticisms are correct, the opposite view is arrived at here, and so, if the present theory is correct, the "anomalous processes" introduced by Peierls have little importance for the electrical conductivity in a constant field.

† 'Z. Physik,' vol. 52, p. 555 (1928).

‡ 'Ann. Physik,' vol. 4, p. 121 (1930).

§ 'Ann. Physik,' vol. 10, p. 97 (1931), and vol. 12, p. 154 (1932).

|| 'Die Quantenstatistik,' chap. 8, Springer (1931).

General Theory.

2. The method adopted is that used in dealing with molecules. In the first approximation the electrons and the nuclei are treated separately, and the interactions are brought in in the second approximation. Care is, however, necessary, as the first approximation when we have particles of widely differing mass is by no means the same as the usual first approximation which arises when we treat, for example, the interaction of the two electrons in a helium atom.

We consider any finite number of electrons and nuclei. The electron co-ordinates are $x_\gamma, y_\gamma, z_\gamma$, while the co-ordinates of the nuclei are X_g, Y_g, Z_g . The Schrödinger equation of the assembly is

$$\left[\frac{1}{m} \sum_{\gamma} \nabla_{xyz, \gamma}^2 + \frac{1}{M} \sum_g \nabla_{XYZ, g}^2 - \frac{8\pi^2}{h^2} V(x, X) - \frac{4\pi i}{h} \frac{\partial}{\partial t} \right] \Psi = 0, \quad (1)$$

where m is the mass of an electron, M the mass of a nucleus, and $V(xX)$ is the total potential energy of the system. Since m/M is small, we can separate the equation approximately as follows. If

$$\left[\frac{1}{m} \sum_{\gamma} \nabla_{\gamma}^2 + \frac{8\pi^2}{h^2} \{E(X) - V(xX)\} \right] \psi(xX) = 0, \quad (2)$$

and

$$\left[\frac{1}{M} \sum_g \nabla_g^2 + \frac{8\pi^2}{h^2} \{W - E(X)\} \right] \phi(X) = 0 \quad (3)$$

have solutions $\psi(xX)$ and $\phi(X)$, then an approximate solution of (1) is

$$\Psi = \psi(xX) \phi(X) e^{2\pi i W t / h}.$$

In equation (2) the nuclear co-ordinates are merely parameters, and the equation is the Schrödinger equation for the electrons, the nuclei being fixed in arbitrary positions. To obtain a more accurate solution, suppose that the solutions of (2) and (3) are $\psi_r(xX)$ and $\phi_r(X)$, the corresponding energy values being $E_r(X)$ and W_r . Then we solve (1) exactly by

$$\Psi = \sum_{rp} a_{rp}(t) \psi_r(xX) \phi_r(X) \exp(2\pi i W_r t / h), \quad (4)$$

the equations determining the coefficients being

$$\dot{a}_{rp}(t) = \sum_{s\sigma} a_{s\sigma}(t) (r\rho | A | s\sigma) \exp 2\pi i t (W_s^{\sigma} - W_r^{\rho}) / h, \quad (5)$$

where

$$(r\rho|A|s\sigma) = \frac{\hbar}{4\pi M i} \sum_{\sigma} \left[\int \phi_{\rho}^r(\mathbf{X})^* \phi_{\sigma}^s(\mathbf{X}) \psi_r(x\mathbf{X})^* \nabla_{\mathbf{X},y}^2 \psi_s(x\mathbf{X}) dx d\mathbf{X} \right. \\ \left. + 2 \int \phi_{\rho}^r(\mathbf{X})^* \psi_r(x\mathbf{X})^* (\text{grad}_{\mathbf{X},y} \phi_{\sigma}^s(\mathbf{X}) \cdot \text{grad}_{\mathbf{X},y} \psi_s(x\mathbf{X})) dx d\mathbf{X} \right]. \quad (6)$$

These equations are exact, but as usual we only proceed to solve them approximately, so that the probability per unit time that the electrons undergo a transition $r \rightarrow s$, while the lattice makes a transition $\rho \rightarrow \sigma$ is

$$|(r\rho|A|s\sigma)|^2 \Omega (W_s^r - W_r^{\rho}),$$

where

$$\Omega(x) = \frac{\partial}{\partial t} \frac{1 - \cos 2\pi x t / \hbar}{x^2}.$$

The question as to whether this approximation is a sufficiently good one is extremely difficult. We are here dealing with the perturbations of the continuous spectrum, and it is not easy to give any criterion for the rapid convergence of the series of approximations, which justifies us in stopping at the second term. The best that we can do seems to be to apply the theory as it stands, and to assume that the procedure is justified so long as it does not give rise to any infinities. This is, of course, the method adopted in dealing with the collisions of electrons and photons with atoms. In order to evaluate the transition probabilities it is necessary to obtain an expression for $\psi(x\mathbf{X})$. This we cannot do in full generality, but provided we assume that the lattice vibrations are of small amplitude we can obtain approximate expressions, which will be valid if the metal is not near the melting point.

Assumptions.

3. It is convenient at this stage to collect together the major assumptions to be made. We suppose that in equilibrium at the absolute zero the lattice is simple cubic with lattice constant a . We number the lattice points by the integers g_1, g_2, g_3 , and we now suppose that $\mathbf{X}_{g_1 g_2 g_3}$ represents the vector displacement of the (g_1, g_2, g_3) nucleus from its equilibrium position. The quantity $E(\mathbf{X})$, which acts as a kind of potential energy for the nuclear vibrations, can be expanded in terms of the \mathbf{X} 's, and we use only the quadratic terms since the linear terms must necessarily be zero. We assume that, for the electronic states which are of importance in conduction, $E_r(\mathbf{X})$ is independent of r . This implies that the lattice constant is not affected by the excitation of

an electron, an assumption which is justified to the first order for a solid in bulk, but which is not true, for example, for a diatomic molecule.

The next assumption we have to make concerns the form of the electronic wave functions for the equilibrium positions of the nuclei. We shall assume that the electrons may be treated independently, and as moving in a periodic field of force, the potential energy of an electron being given as a sum over the lattice points

$$\begin{aligned} V(x, 0) &= \sum_{g_1, g_2, g_3 = -\infty}^{\infty} U(x - g_1 a, y - g_2 a, z - g_3 a) \\ &= \sum_{\mathbf{g} = -\infty}^{\infty} U(\mathbf{r} - \mathbf{g}a). \end{aligned} \quad (7)$$

We shall also assume that the electrons are moderately tightly bound, and that the wave functions of the lowest states are

$$\psi_{klm}(x, 0) = \sum_{g_1, g_2, g_3 = -\infty}^{\infty} e^{\frac{2\pi i}{G}(kg_1 + lg_2 + mg_3)} \phi(\mathbf{r} - \mathbf{g}a), \quad (8)$$

where $\phi(\mathbf{r} - \mathbf{g}a) = \phi_{g_1, g_2, g_3}$ is a function of distance only, and is only large in the neighbourhood of the lattice point (g_1, g_2, g_3) . We are here using the usual periodic boundary condition, that all the properties of the crystal are periodic in a large cube of side Ga . The actual form of the electronic wave functions is not very important in the present state of the theory, so long as the electrons are not taken to be entirely free. This is because the Fermi function is very insensitive to changes in temperature, and any phenomena in a metal which show large temperature variations must be conditioned mainly by the nuclear vibrations, the form of the electronic wave functions only playing a secondary part. Phenomena which are nearly independent of temperature lie completely outside the scope of the theory. We shall see later how far these various assumptions are consistent with each other.

The Motion of the Lattice.

4. The problem here is to reduce the potential energy to the sum of squares, and we therefore consider, for the moment, the problem of the motion of a lattice in classical mechanics. The normal modes of a lattice have been investigated by Born,[†] but the only complete theory is that given by Waller,[‡]

[†] 'Atomtheorie des Festen Zustandes,' Teubner (1923).

[‡] 'Uppsala Arsskrift,' vol. 2 (1925).

who shows that the normal co-ordinates given by Born are not all independent. If we write

$$\mathbf{X}_{fg_h} = \sum'_{fgh=-\frac{1}{2}G}^{\frac{1}{2}G} \sum_{j=1}^3 \left[a_{fghj} \cos \frac{2\pi}{G} (fg_1 + gg_2 + hg_3) + b_{fghj} \sin \frac{2\pi}{G} (fg_1 + gg_2 + hg_3) \right] \mathbf{u}_{fghj} \quad (9)$$

then in the limit as $G \rightarrow \infty$, the a 's and b 's become the normal co-ordinates, while for finite values of G they differ from the correct normal co-ordinates by quantities of the order $1/G^3$. Here \mathbf{u}_{fghj} is a unit vector, which for $j = 1$ is parallel to the direction (f, g, h) (longitudinal waves) and for $j = 2, 3$ is perpendicular to (f, g, h) (transverse waves). The meaning of Σ' is as follows. Draw any plane through the centre of the cube $-\frac{1}{2}G \leq f, g, h \leq \frac{1}{2}G$, which is called the phase cube. Then Σ' denotes a summation over the points lying in either half of the cube, giving the correct number $(3G^3)$ of normal modes. It should be noted that the above expression does not limit the waves to be standing waves, since the coefficients may be chosen so as to give either standing waves or progressive waves.

It is at this point that we differ from Peierls. He writes, for a one-dimensional lattice,

$$X_{g_1} = \sum_{f=-\frac{1}{2}G}^{\frac{1}{2}G} \xi_f e^{2\pi i f g_1 / G},$$

and treats the ξ_f as the normal co-ordinates. But X_{g_1} , being the displacement of a material point, is necessarily real and so we must have $\xi_f^* = \xi_{-f}$. We must therefore take the real and imaginary parts of ξ_f as the independent quantities, and, writing $\xi_f = \frac{1}{2}(a_f - ib_f)$, we obtain

$$X_{g_1} = \sum_{f=0}^{\frac{1}{2}G} \left[a_f \cos \frac{2\pi}{G} f g_1 + b_f \sin \frac{2\pi}{G} f g_1 \right],$$

which is the one-dimensional form of (9).

The equations satisfied by the normal co-ordinates are

$$\begin{aligned} \ddot{a}_{fghj} + 4\pi^2 \nu_{fghj}^2 a_{fghj} &= 0, \\ \ddot{b}_{fghj} + 4\pi^2 \nu_{fghj}^2 b_{fghj} &= 0. \end{aligned}$$

Here ν_{fghj} is a complicated function of the coefficients in the quadratic form $E(\mathbf{X})$, but for long waves, that is for small values of f, g, h , we may assume that the lattice behaves as a continuum, and we then have

$$\nu_{fghj} = \frac{c_j}{Ga} \sqrt{(f^2 + g^2 + h^2)}, \quad (10)$$

where c_j is the velocity of sound, supposed constant, for each type of vibration. With these co-ordinates the energy of the lattice vibrations becomes

$$\begin{aligned} H &= \frac{1}{2} (\frac{1}{2} MG^3) \sum'_{fgh} \sum_{j=1}^3 [\dot{a}_{fghj}^2 + \dot{b}_{fghj}^2 + 4\pi^2 v_{fghj}^2 (a_{fghj}^2 + b_{fghj}^2)] \\ &= \frac{1}{2} \sum'_{fgh} \sum_{j=1}^3 \left[\frac{2}{MG^3} (p_a^2 + p_b^2) + (\frac{1}{2} MG^3) 4\pi^2 v_{fghj}^2 (a_{fghj}^2 + b_{fghj}^2) \right]. \quad (11) \end{aligned}$$

It is now a simple matter to solve the Schrödinger equation (3) for the lattice, since it is separable in the co-ordinates a_f , b_f , and the corresponding momenta.

We therefore write

$$\phi(X) = \prod'_{fghj} \phi(a_{fghj}) \phi(b_{fghj}), \quad (12)$$

where $\phi(a)$ satisfies the equation for a simple harmonic oscillator

$$\frac{d^2\phi}{da_{fghj}^2} + \frac{8\pi^2 M_0}{h^2} \{W - 2\pi^2 M_0 v_{fghj}^2 a_{fghj}^2\} \phi = 0, \quad (13)$$

M_0 being written for $\frac{1}{2} MG^3$. On transforming to the new variables, the expression (6), which determines the interaction between the electrons and the lattice, is unaltered in form, differentiations with respect to X_g being replaced by differentiations with respect to a_f and b_f , and M being replaced by M_0 .

The Electronic Wave Functions.

5.1. We have now to determine the electronic wave functions as functions of the nuclear co-ordinates. This we do by perturbation theory, assuming that the displacements of the lattice are small compared with the lattice constant. The wave function $\psi(xX)$ of a single electron satisfies the equation

$$\left[\frac{1}{2m} \nabla_{xyz}^2 + \frac{8\pi^2}{h^2} \{E(X) - V(xX)\} \right] \psi(xX) = 0, \quad (14)$$

and we suppose that we can solve this equation when the nuclei are at their equilibrium positions, the wave function being then $\psi(x, 0)$, satisfying

$$\left[\frac{1}{2m} \nabla_{xyz}^2 + \frac{8\pi^2}{h^2} \{E(0) - V(x, 0)\} \right] \psi(x, 0) = 0. \quad (15)$$

We consider (14) as a perturbation on (15), the perturbing energy being

$$\Delta V = V(xX) - V(x, 0).$$

Now

$$V(xX) = \sum_{g_1 g_2 g_3 = -\infty}^{\infty} U(x - g_1 a - X_{g_1 g_2 g_3}, y - g_2 a - Y_{g_1 g_2 g_3}, z - g_3 a - Z_{g_1 g_2 g_3}),$$

and so

$$\Delta V = - \sum_{g_1 g_2 g_3 = -\infty}^{\infty} (\mathbf{X}_{g_1 g_2 g_3} \cdot \text{grad } U(\mathbf{r} - \mathbf{g}a)). \quad (16)$$

To calculate the matrix elements of ΔV with respect to the solutions of (15), which we take to be given by (8), we write ΔV as

$$- \sum_{g_1 g_2 g_3 = -\infty}^{\infty} \sum'_{fgh} \sum_{j=1}^3 \left[\frac{1}{2} (a_f - ib_f) \exp \frac{2\pi i}{G} (fg_1 + gg_2 + hg_3) \right. \\ \left. + \frac{1}{2} (a_f + ib_f) \exp - \frac{2\pi i}{G} (fg_1 + gg_2 + hg_3) \right] (\mathbf{u}_{fgh} \cdot \text{grad } U_{g_1 g_2 g_3}). \quad (17)$$

In the expression for the matrix element $\langle k'l'm' | \Delta V | klm \rangle$, the terms arising from the positive exponential have as factor

$$\int \psi^*_{k'l'm'} \sum_{g_1 g_2 g_3} e^{\frac{2\pi i}{G} (fg_1 + gg_2 + hg_3)} (\mathbf{u}_{fgh} \cdot \text{grad } U_{g_1 g_2 g_3}) \psi_{klm} dx dy dz, \quad (18)$$

where the integration is over the large cube of side Ga . Now

$$\psi_{klm} = \sum_{p_1 p_2 p_3 = -\infty}^{\infty} e^{\frac{2\pi i}{G} (kp_1 + lp_2 + mp_3)} \phi_{p_1 p_2 p_3}$$

and

$$\psi^*_{k'l'm'} = \sum_{p'_1 p'_2 p'_3 = -\infty}^{\infty} e^{-\frac{2\pi i}{G} (k'p'_1 + l'p'_2 + m'p'_3)} \phi_{p'_1 p'_2 p'_3},$$

so there are nine summations to be carried out in (18). Most of the terms are, however, zero, when we use the approximations which have already been made in deriving (8), namely, that the functions ϕ are only appreciable in the neighbourhood of one lattice point and its immediate neighbours. Therefore of the terms in (18) involving $\partial U_{g_1 g_2 g_3} / \partial x$ we retain only those for which

$$p_1 = p'_1 = g_1 \pm 1$$

$$p_2 = p'_2 = g_2$$

$$p_3 = p'_3 = g_3,$$

the terms with $(g_1 + 1)$ being equal in magnitude but opposite in sign to those with $(g_1 - 1)$. There are corresponding terms arising from the other components of the gradient. Writing

$$\int \frac{\partial U_{g_1 g_2 g_3}}{\partial x} \phi^2_{g_1+1, g_2, g_3} dx dy dz = C$$

if $(g_1 g_2 g_3)$ lies in the large cube, and putting this expression equal to zero otherwise, we see that (18) reduces to

$$2iC \sum_{f_1 f_2 f_3 = -\frac{1}{2}G}^{\frac{1}{2}G} \left(\mathbf{u}_{f g h j} \cdot \sin \frac{2\pi}{G} (\mathbf{k} - \mathbf{k}') \right) \exp \frac{2\pi i}{G} \{ (k - k' + f) g_1 + (l - l' + g) g_2 + (m - m' + h) g_3 \},$$

which is equal to

$$- 2iCG^3 \left(\mathbf{u}_{f g h j} \cdot \sin \frac{2\pi}{G} \mathbf{f} \right) \quad (19)$$

if

$$\mathbf{k} - \mathbf{k}' + \mathbf{f} = 0 \pmod{G}, \quad (20)$$

and is zero otherwise. Therefore for small values of f, g, h , the transverse waves give zero contribution, and we shall in future omit the suffix j and deal only with the longitudinal vibrations. There are corresponding terms arising from the negative exponential in (17), which are zero unless

$$\mathbf{k} - \mathbf{k}' - \mathbf{f} = 0 \pmod{G}. \quad (21)$$

Those transitions for which $k - k' \pm f = G$ have been called anomalous transitions by Peierls, and he maintains that they are the terms which dominate the electrical conduction. The reason he gives is that the condition (20), if we only allow 0 on the right-hand side, is a conservation theorem for the impulse *quantum numbers*, and if we only consider these transitions the conductivity is infinite. This is quite correct, but since we are not limited to the transitions (20) and have in addition the transitions (21), the argument breaks down, and the conductivity is finite when we omit the anomalous transitions. Of course the transitions (21) occur in Peierls' theory, but, since his normal co-ordinates are not independent, the transitions (20) and (21) arise from different normal co-ordinates and so can never in any way compensate one another. In future we shall therefore neglect the anomalous transitions, and the rest of this paper is concerned with the justification of this procedure, which consists in showing that the conductivity is finite when we do neglect them. In passing we may notice that (19) confirms Peierls' formulæ as against those of Brillouin, who wrote $2\pi(\mathbf{k} - \mathbf{k}')/G$ instead of $\sin 2\pi(\mathbf{k} - \mathbf{k}')/G$. We finally have, on collecting terms together and inserting the proper normalising factors,

$$\begin{aligned} \psi_{klm}(xX) = \psi_{klm}(x, 0) + \sum_{f \neq 0} \left[\frac{(k+f|\Delta V|k)}{E_k - E_{k+f}} \psi_{k+f}(x, 0) \right. \\ \left. + \frac{(k-f|\Delta V|k)}{E_k - E_{k-f}} \psi_{k-f}(x, 0) \right], \quad (22) \end{aligned}$$

where

$$\left. \begin{aligned} (k+f|\Delta V|k) &= -\frac{2\pi i C}{G} \sqrt{(f^2 + g^2 + h^2)} (a_f - ib_f) \\ \text{and} \\ (k-f|\Delta V|k) &= \frac{2\pi i C}{G} \sqrt{(f^2 + g^2 + h^2)} (a_f + ib_f). \end{aligned} \right\} \quad (23)$$

5.2. We are now in a position to discuss the compatibility of some of the assumptions we have made. The energy $E(X)$ of a single electron is given to the second order by

$$\begin{aligned} E_k(X) &= E_k(0) + \sum'_{fgh} \left[\frac{|(k+f|\Delta V|k)|^2}{E_k - E_{k+f}} + \frac{|(k-f|\Delta V|k)|^2}{E_k - E_{k-f}} \right] \\ &= E_k(0) + \frac{4\pi^2 C^2}{G^2} \sum'_{fgh} \left[(f^2 + g^2 + h^2) (a_f^2 + b_f^2) \left(\frac{1}{E_k - E_{k+f}} + \frac{1}{E_k - E_{k-f}} \right) \right]. \end{aligned}$$

The energy is therefore expressed as the sum of squares, as it ought to be. To compare this with (11) it is necessary to sum over all the electrons, and this we cannot do without a knowledge of the form of E_k . However, it is obvious that on account of the occurrence of the quantities $E_{k \pm f}$, ν_{fgh} will not be given exactly by equation (10), but we shall not take this refinement into account.

The Interaction.

6. We have now to introduce the interaction, which is given by equations (5) and (6). To the order to which we are working, the first set of terms is zero, since it involves second derivatives of $\psi(xX)$. In a more exact theory these terms would have to be included, but we shall omit them here. The expression (6) therefore becomes

$$(r\rho|A|s\sigma) = \frac{h}{2\pi M_0 i} \sum'_{fgh} \int \phi_\rho(X)^* \psi_r(xX)^* \left\{ \frac{\partial \phi_\sigma}{\partial a_f} \frac{\partial \psi_s}{\partial a_f} + \frac{\partial \phi_\sigma}{\partial b_f} \frac{\partial \psi_s}{\partial b_f} \right\} dx dX. \quad (24)$$

In this we may write $\psi_r(x, 0)^*$ instead of $\psi_r(xX)^*$, and from (22) it is obvious that the only electronic transitions possible are those for which an electron jumps from the state k to the state $k \pm f$, using the notation of the last section. Further, since $\phi(X)$ is a product of functions of the individual normal coordinates, only one of the oscillators can make a transition, namely, the f th oscillator. As usual, the only transitions of the oscillator which are allowed

are those in which the quantum number changes by unity. To show this we have to evaluate

$$\int_{-\infty}^{\infty} \phi_{\rho f}'(a_{fg\hbar}) \frac{d}{da_{fg\hbar}} \phi_{\sigma f}(a_{fg\hbar}) da_{fg\hbar}. \quad (25)$$

Now the unnormalised solutions of (13) are

$$\phi_{\sigma}(a) = e^{-\frac{1}{2}\lambda^2 a^2} H(\lambda a)$$

and

$$W_{\sigma} = \hbar \nu_{fg\hbar} (\sigma + \frac{1}{2}), \quad (26)$$

where H_{σ} is the σ th Hermite polynomial, and

$$\lambda^2 = 4\pi^2 \nu_f M_0 / \hbar. \quad (27)$$

Using the recurrence relations for the Hermite polynomials we find

$$\frac{d\phi_{\sigma}}{da} = \lambda \sigma \phi_{\sigma-1} - \frac{1}{2}\lambda \phi_{\sigma+1},$$

and so the integral (25) vanishes unless $|\sigma - \rho| = 1$. Putting in the proper normalising factors, the only non-vanishing integrals are

$$\int_{-\infty}^{\infty} \phi_{\sigma \pm 1}(a) \frac{d}{da} \phi_{\sigma}(a) da = -\left\{\frac{1}{2}\lambda^2 (\sigma + 1)\right\}^{\frac{1}{2}} \left(\frac{1}{2}\lambda^2 \sigma\right)^{\frac{1}{2}}.$$

Hence the only transitions possible are those in which the f th oscillator emits or absorbs a quantum $\hbar \nu_f$, while an electron changes its quantum numbers from \mathbf{k} to \mathbf{k}' , where $\mathbf{k} - \mathbf{k}' \pm \mathbf{f} = 0$. The probabilities per unit time of these transitions occurring are

$$|(k, \sigma_f | A | k \pm f, \sigma_f - 1)|^2 \Omega (E_k - E_{k'} + \hbar \nu_f)$$

and

$$|(k, \sigma_f | A | k \pm f, \sigma_f + 1)|^2 \Omega (E_k - E_{k'} - \hbar \nu_f), \quad (28)$$

where

$$|(k, \sigma_f | A | k', \sigma_f \pm 1)|^2 = \frac{\hbar C^2}{M_0 (E_k - E_{k'})^2} \frac{\nu_f}{G^2} (f^2 + g^2 + \hbar^2) \begin{Bmatrix} \sigma_f + 1 \\ \sigma_f \end{Bmatrix}. \quad (29)$$

In deriving this result we have used the fact that the b oscillators give the same contribution as the a oscillators.

The Conduction Problem.

7.1. We now apply the microscopic transition probabilities to the statistical conduction theory. In the first place, owing to the distribution of energy

between the various states of an oscillator, we have to use an average value of σ_f . In thermal equilibrium we have the usual formula for simple harmonic oscillators

$$\bar{\sigma}_f = \frac{1}{e^{\hbar\nu_f/kT} - 1}.$$

We shall not assume that, when a current is flowing, $\bar{\sigma}_f$ is given by this expression, and we shall write instead N_f . The electronic distribution function we shall call $n(k)$, which, in the absence of external fields, is the Fermi function. Then, if we write

$$A^{\pm}_{kk'f} = \frac{\hbar C^2}{M_0} \frac{\nu_f}{(E_k - E_{k'})^2} \frac{4\pi^2 (f^2 + g^2 + h^2)}{G^2} \Omega(E_k - E_{k'} \pm \hbar\nu_f), \quad (27)$$

the number of electrons which are forced into the state k by collisions, per unit time, is

$$\sum_{k'f} n(k') \{1 - n(k)\} A^+_{kk'f} (N_f + 1) + \sum_{k'f} n(k') \{1 - n(k)\} A^-_{kk'f} N_f, \quad (28)$$

and the number ejected from k by collisions per unit time is

$$\sum_{k'f} n(k) \{1 - n(k')\} A^+_{kk'f} N_f + \sum_{k'f} n(k) \{1 - n(k')\} A^-_{kk'f} (N_f + 1), \quad (29)$$

where

$$k - k' \pm f = 0.$$

Further, the net number of quanta of energy $\hbar\nu_f$ created is

$$\begin{aligned} & \sum_{kk'} [n(k') \{1 - n(k)\} (N_f + 1) - n(k) \{1 - n(k')\} N_f] A^+_{kk'f} \\ & + \sum_{kk'} [-n(k') \{1 - n(k)\} N_f + n(k) \{1 - n(k')\} (N_f + 1)] A^-_{kk'f}. \end{aligned} \quad (30)$$

Now we are interested in the steady state set up by an electric field when the metal is maintained at constant temperature. Therefore the number of electrons lost to the state k by collisions, which is the difference between (28) and (29), must be balanced by the number restored by the field, which, if the field is F parallel to the axis of x , is

$$\frac{GeaF}{\hbar} \frac{\partial n_0}{\partial k}. \quad (31)$$

Also since the lattice is in a steady state, the net number of quanta created must be zero, that is (30) is zero.

7.2. To solve the above equations we put

$$n(k) = n_0(k) + F n_1(k) \frac{1}{(e^\varepsilon + 1)(e^{-\varepsilon} + 1)} \quad (32)$$

and

$$N_f = N_f^0 + F N_f^{(1)} \frac{1}{(e^x - 1)(e^{-x} + 1)}, \quad (33)$$

where

$$n_0(k) = \frac{1}{e^{(\varepsilon - \varepsilon_0)/kT} + 1} = \frac{1}{e^{\varepsilon(k)} + 1}, \quad (34)$$

and

$$N_f^0 = \frac{1}{e^{h\nu_f/kT} - 1} = \frac{1}{e^{x(f)} - 1}. \quad (35)$$

The quantity $N_f^{(1)}$ is usually assumed to be zero, but the crux of Peierls' theory is that it does not vanish. This we shall show to be wrong. Neglecting squares of F we have

$$\begin{aligned} \frac{Gea}{h} \frac{\partial n_0}{\partial k} = & \sum'_{fgh} \frac{A^+_{k, k+f, f}}{\{e^{\varepsilon(k)} + 1\} \{e^{\varepsilon(k+f)} + 1\} \{e^{x(f)} - 1\}} [n_1(k+f) - n_1(k) - N_f^{(1)}] \\ & + \sum'_{fgh} \frac{A^+_{k, k-f, f}}{\{e^{\varepsilon(k)} + 1\} \{e^{\varepsilon(k-f)} + 1\} \{e^{x(f)} - 1\}} [n_1(k-f) - n_1(k) - N_f^{(1)}] + \dots \end{aligned} \quad (36)$$

For convenience we have not written out the corresponding terms involving A^- , and we have also used the fact that in any expression multiplying $A^+_{kk'f}$ we may put $\varepsilon - \varepsilon' + x = 0$ with negligible error.

Further

$$\begin{aligned} \sum_{klm} \frac{A^+_{k, k+f, f}}{\{e^{\varepsilon(k)} + 1\} \{e^{\varepsilon(k+f)} + 1\} \{e^{x(f)} - 1\}} [n_1(k+f) - n_1(k) - N_f^{(1)}] \\ + \sum_{klm} \frac{A^+_{k, k-f, f}}{\{e^{\varepsilon(k)} + 1\} \{e^{\varepsilon(k-f)} + 1\} \{e^{x(f)} - 1\}} [n_1(k-f) - n_1(k) - N_f^{(1)}] + \dots = 0, \end{aligned} \quad (37)$$

where we have again omitted the terms in A^- . The summation with respect to f, g, h is over half the phase cube, as explained above, while the summation with respect to k, l, m is over the cube $-\frac{1}{2}G \leq k, l, m \leq \frac{1}{2}G$. The expressions (36) and (37) may be immediately simplified on symmetry grounds. The function $n_1(k)$ is an odd function of k , since the change in the electronic distribution function is conditioned by $\partial n_0/\partial k$, and therefore

$$n_1(k, l, m) = -n_1(-k, -l, -m).$$

On the other hand $\varepsilon(k)$ and $A_{kk'f}$ are both even functions of k, l, m , provided

$E(k)$ is an even function, which is always the case. Therefore in the second term in (37) change k into $-k$. Then $A^+_{k, k-f, f}$ becomes $A^+_{k, k+f, f}$ and $n_1(k-f)$ and $n_1(k)$ become $-n_1(k+f)$ and $-n_1(k)$ respectively. The terms involving n_1 disappear completely, and so

$$N_f^{(1)} = 0.$$

It is not difficult to see why Peierls was led to a different result. As explained in Section 4, he omitted the relation between his normal co-ordinates, $\xi_f^* = \xi_{-f}$, and so the second term in (37) is missing, as the first and second terms would then arise from different normal co-ordinates. A solution of (37) would then be $n_1(k) = k$ and $N_f^{(1)} = f$, and this makes the conductivity infinite, as shown by Peierls. However, the only correct solution is $N_f^{(1)} = 0$, and the lattice is automatically in thermal equilibrium if the electrons are in a steady state.

Now in the second term in (36) change f into $-f$. We then obtain exactly the first term, but summed over the other half of the phase cube. Thus finally (36) becomes

$$\begin{aligned} \frac{Gea}{h} \frac{\partial n_0}{\partial k} = & \sum_{f/h = -\frac{1}{2}G}^{\frac{1}{2}G} \frac{A^+_{k, k+f, f}}{\{e^{\epsilon(k)} + 1\} \{e^{\epsilon(k+f)} + 1\} \{e^{\epsilon(f)} - 1\}} [n_1(k+f) - n_1(k)] \\ & + \sum_{f/h = -\frac{1}{2}G}^{\frac{1}{2}G} \frac{A^-_{k, k+f, f}}{\{e^{\epsilon(k)} + 1\} \{e^{\epsilon(k+f)} + 1\} \{e^{\epsilon(f)} - 1\}} [n_1(k+f) - n_1(k)]. \quad (38) \end{aligned}$$

This is effectively the same as Bloch's result, but with different meanings for some of the constants involved. The subsequent work is now identical with that of Bloch and leads to a conductivity which varies as T^{-1} for high temperatures, and as T^{-5} for low temperatures.

Summary.

A method is given for dealing with the interaction between the electronic motions and the nuclear vibrations in a metal, which is analogous to that used for diatomic molecules. This method is then used to evaluate the "mean free path" of the electrons, and the existing theories are critically discussed. The conclusion arrived at is that the refinements introduced by Peierls are unfounded, and the theory is thereby greatly simplified.

The Torsion and Flexure of Shafting with Keyways or Cracks.

By W. M. SHEPHERD.

(Communicated by L. N. G. Filon, F.R.S.—Received August 2, 1932.)

1. *The Object of the Paper.*

The object of the paper is to investigate the properties of shafts of circular cross-section into which keyways or slits have been cut, first when subjected to torsion, and second when bent by a transverse load at one end. The torsion problem for similar cases has been treated by several writers. Filon* has worked out an approximation to the case of a circular section with one or two keyways; in his method the boundary of the cross-section was a nearly circular ellipse and the boundaries of the keyways were confocal hyperbolas. In particular he considered the case when the hyperbola degenerated into straight lines starting from the foci. The solution for a circular section with one keyway in the form of an orthogonal circle has been obtained by Gronwall.† In each case the solution has been obtained by the use of a conformal transformation and this method is again used in this paper, the transformations used being

$$\rho = k \operatorname{sn}^2 t, \quad \rho = k^{\frac{1}{2}} \operatorname{sn} t, \quad \rho = k^{\frac{1}{2}} \operatorname{sn}^{\frac{1}{2}} t$$

where

$$\rho = x + iy, \quad t = \xi + i\eta.$$

No work appears to have been done on the flexure problem which is here worked out for several cases of shafts with slits.

2. *Summary of the Problems Treated.*

We first consider the torsional properties of shafts with one and with two indentations. In particular cases numerical results have been obtained for the stresses at particular points and for the torsional rigidity. The results for one indentation and for two indentations of the same width and approximately the same depth have been compared. We next consider the solution of the torsion problem for one, two or four equal slits of any depth from the surface towards the axis. The values of the stresses have not been worked out in these cases since the stress is infinite at the bottom of the slits. This infinite

* 'Phil. Trans.,' A, vol. 193, pp. 309–352 (1900).

† 'Trans. Amer. Math. Soc,' vol. 20, p. 234 (1919).

stress occurs because the physical conditions are not satisfied at the bottom of the slits, but as had been pointed out by Filon* this does not affect the validity of the values of the torsional rigidity. We compare the effect on the torsional rigidity of the shaft of one, two and four slits of the same depth in particular cases. We also compare the results for one slit with those obtained by Filon* by another method, and find very good agreement which is illustrated by a graph. The reduction in torsional rigidity due to a semicircular keyway is compared with that due to a slit of approximately the same depth. Finally the distortion of the cross-sections at right angles to the planes is investigated, and in this, several interesting and perhaps unexpected features appear. The relative shift of the two sides of the slits is calculated in several cases.

The flexure problem is then solved for circular shafts with one or two slits as in the case of torsion. The methods adopted are very similar to those of the first part of the paper. The distortion of the cross-sections and the relative shift at the lips of the slits is again investigated.

3. *Torsion—Statement of the Notation.*

We take the generators of the shaft (assumed of uniform cross-section) to be parallel to the axis of z .

Then the displacements are given by†

$$u = -\tau yz, \quad v = \tau zx, \quad w = \tau \phi, \quad (3.1)$$

where ϕ is a function of x and y and τ is the twist.

The stresses are

$$\begin{aligned} \widehat{xx} = \widehat{yy} = \widehat{zz} = \widehat{xy} &= 0, \\ \widehat{xz} = \mu\tau \left[\frac{\partial \phi}{\partial x} - y \right], \quad \widehat{yz} = \mu\tau \left[\frac{\partial \phi}{\partial y} + x \right], \end{aligned} \quad (3.11)$$

and the function ϕ satisfies the equation

$$\nabla^2 \phi \equiv \frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} = 0,$$

within the boundary of the cross-section and certain conditions on the boundary.

The conjugate function ψ , defined by

$$\frac{\partial \phi}{\partial x} = \frac{\partial \psi}{\partial y}, \quad \frac{\partial \phi}{\partial y} = -\frac{\partial \psi}{\partial x}, \quad (3.12)$$

* *Loc. cit.*, p. 350.

† Love, "Mathematical Theory of Elasticity," chap. 14 (1927).

satisfies

$$\nabla^2 \psi = 0$$

within the boundary, and on the boundary the condition

$$\psi = \frac{1}{2}(x^2 + y^2) + \text{constant.} \quad (3.13)$$

The torsion problem consists of finding a plane harmonic function inside the boundary which satisfies (3.13) on the boundary.

4. The Transformations Employed.

(i) Consider the transformation*

$$\rho = x + iy = k \operatorname{sn}^2 t = k \operatorname{sn}^2 (\xi + i\eta). \quad (4.1)$$

When

$$\eta = \pm \frac{1}{2}K', \quad r = |\rho| = 1,$$

and when

$$\xi = \pm \frac{1}{2}K', \quad r = |\rho| = \frac{k}{1+k'} \frac{1+k' \operatorname{sn}^2(\eta, k')}{1-k' \operatorname{sn}^2(\eta, k')}, \quad (4.21)$$

$$= \{\operatorname{dn}(2\eta, k') - k' \operatorname{cn}(2\eta, k')\}/k. \quad (4.22)$$

If $y = x \tan \theta$, then for points on $\xi = \pm \frac{1}{2}K$, $-\frac{1}{2}K' < \eta < \frac{1}{2}K'$,

$$\tan \frac{1}{2}\theta = k' \operatorname{sn}(\eta, k') \operatorname{cn}(\eta, k')/\operatorname{dn}(\eta, k'). \quad (4.23)$$

The relation between r and θ obtained by eliminating η between (4.21) and (4.23) is

$$\cos \theta = k(1+r^2)/2r.$$

It follows that the rectangle $\xi = \pm \frac{1}{2}K$, $\eta = \pm \frac{1}{2}K'$ corresponds to the unit circle with one indentation in the form of an orthogonal circle, the common points being at $\rho = k \pm ik'$, the indented circle being taken twice in passing round the rectangle. The boundary of the indentation crosses the axis of x at $x = k/(1+k') = (1-k')/k$ (see fig. 1).

The rectangle $\xi = \pm K$, $\eta = \pm \frac{1}{2}K'$ corresponds in the same way to the unit circle with one slit from $x = k$ to $x = 1$. When

$$\xi = \pm K, \quad r = |\rho| = x = k \operatorname{nd}^2(\eta k'). \quad (4.3)$$

The value of ρ is unchanged if we change the sign of t , thus giving the one-two correspondence of the planes of ρ and t . Since the functions used in the

* Cf. Forsyth, "Theory of Functions," p. 501 (1893), and Cayley, 'Trans. Camb. Phil. Soc.,' vol. 14, p. 484 (1889).

application of this transformation must necessarily be one valued functions of x and y they must be such as to be unaltered when t changes sign. It follows that the functions must be even in both ξ and η or odd in both ξ and η . It will be seen that these conditions are satisfied in every case in which this transformation is employed.

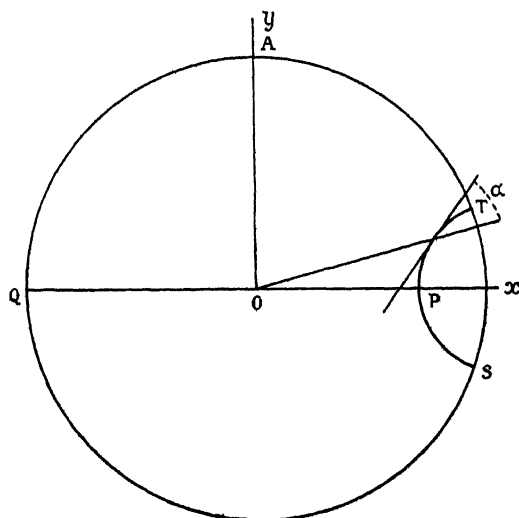


FIG. 1.

If we take the boundary values of ψ to be given by

$$2\psi = x^2 + y^2 - 1$$

we get, in the case of one indentation

$$\eta = \pm \frac{1}{2}K', \quad 2\psi = 0, \quad (4.41)$$

$$\xi = \pm \frac{1}{2}K, \quad 2\psi = \{k' \operatorname{cn}^2(2\eta, k') - \operatorname{cn}(2\eta, k') \operatorname{dn}(2\eta, k')\} 2k'/k^2, \quad (4.42)$$

and in the case of one slit

$$\eta = \pm \frac{1}{2}K', \quad 2\psi = 0, \quad (4.51)$$

$$\xi = \pm K, \quad 2\psi = k^2 \operatorname{nd}^4(\eta, k') - 1. \quad (4.52)$$

The evaluation of ψ satisfying these conditions is discussed later.

(ii) Consider the transformation

$$\rho = x + iy = k^{\frac{1}{2}} \operatorname{sn} t = k^{\frac{1}{2}} \operatorname{sn}(\xi + i\eta). \quad (4.6)$$

The rectangle $\eta = \pm \frac{1}{2}K'$, $\xi = \pm \frac{1}{2}K$ now corresponds to the unit circle with two equal indentations each symmetrical about the real axis.

The relation between r and θ is in this case

$$\cos 2\theta = k(1 + r^4)/2r^2. \quad (4.7)$$

The rectangle $\eta = \pm \frac{1}{2}K'$, $\xi = \pm K$ corresponds to the unit circle with two slits from $k^{\frac{1}{2}}$ to 1 and from -1 to $-k^{\frac{1}{2}}$.

In the case of the two indentations the boundary conditions for ψ are

$$\eta = \pm \frac{1}{2}K', \quad 2\psi = 0, \quad (4.81)$$

$$\xi = \pm \frac{1}{2}K, \quad 2\psi = \frac{1}{k} \{ \operatorname{dn}(2\eta, k') - k' \operatorname{cn}(2\eta, k') \} - 1. \quad (4.82)$$

In the case of two slits the conditions are

$$\eta = \pm \frac{1}{2}K', \quad 2\psi = 0, \quad (4.83)$$

$$\xi = \pm K, \quad 2\psi = k \operatorname{nd}^2(\eta, k') - 1. \quad (4.84)$$

(iii) Consider the transformation

$$\rho = x + iy = k^{\frac{1}{2}} \operatorname{sn}^{\frac{1}{2}} t = k^{\frac{1}{2}} \operatorname{sn}^{\frac{1}{2}} (\xi + i\eta).$$

The only problem which can be conveniently solved by this transformation is that of four slits of depth $1 - k^{\frac{1}{2}}$, and the boundary conditions then become

$$\eta = \pm \frac{1}{2}K', \quad 2\psi = 0, \quad (4.91)$$

$$\xi = \pm K, \quad 2\psi = k^{\frac{1}{2}} \operatorname{nd}(\eta, k') - 1. \quad (4.92)$$

5. One Indentation.

The conditions (4.41) and (4.42) can be satisfied by a function of the form

$$\psi = \sum_{m=0}^{\infty} a_{2m+1} \cos [(2m+1)\pi\eta/K'] \cosh [(2m+1)\pi\xi/K'] \operatorname{sech} [(2m+1)\pi K/2K']. \quad (5.1)$$

This satisfies (4.41) and will satisfy (4.42) if we evaluate the a_{2m+1} so that

$$\sum_{m=0}^{\infty} a_{2m+1} \cos [(2m+1)\pi\eta/K'] = [k' \operatorname{cn}^2(2\eta, k') - \operatorname{cn}(2\eta, k') \operatorname{dn}(2\eta, k')] k'/k^2 \quad (5.11)$$

when

$$-\frac{1}{2}K' < \eta < \frac{1}{2}K'.$$

This is possible since the right-hand side is a continuous even function and is zero when $\eta = \pm \frac{1}{2}K'$.

We have*

$$\operatorname{sn} u = \frac{2\pi}{Kk} \left\{ \frac{q^{1/2} \sin x}{1-q} + \frac{q^{3/2} \sin 3x}{1-q^3} + \frac{q^{5/2} \sin 5x}{1-q^5} + \dots \right\},$$

where $u = 2Kx/\pi$ and $q = e^{-\pi k'/k}$.

On differentiating with respect to x we get

$$\operatorname{cn} u \operatorname{dn} u = \frac{\pi^2}{K^2 k} \left\{ \frac{q^{1/2} \cos x}{1-q} + \frac{q^{3/2} \cos 3x}{1-q^3} + \frac{q^{5/2} \cos 5x}{1-q^5} + \dots \right\},$$

and consequently

$$\operatorname{cn}(2\eta, k') \operatorname{dn}(2\eta, k') = \frac{\pi^2}{K'^2 k'} \left\{ \frac{q'^{1/2} \cos(\eta\pi/K')}{1-q'} + \frac{q'^{3/2} \cos(3\eta\pi/K')}{1-q'^3} + \dots \right\}, \quad (5.12)$$

where

$$q' = e^{-\pi k/k'}.$$

This expansion is certainly valid for all values of η which are required. There is no expansion of this form of $\operatorname{cn}^2(2\eta, k')$ valid for all real values of η , so that we must construct a series valid within the range $-\frac{1}{2}K' < \eta < \frac{1}{2}K'$, which is all that we require.

Assume that within this range

$$\operatorname{cn}^2(2\eta, k') = \sum_{m=0}^{\infty} b_{2m+1} \cos[(2m+1)\pi\eta/K']. \quad (5.13)$$

When $\frac{1}{2}K' < \eta < K'$ or $-K' < \eta < -\frac{1}{2}K'$ the right-hand side of (5.13) represents the function $\{-\operatorname{cn}^2(2\eta, k')\}$, and so it follows that

$$\begin{aligned} \frac{1}{2}K' b_{2m+1} &= \int_0^{\frac{1}{2}K'} \operatorname{cn}^2(2\eta, k') \cos[(2m+1)\eta\pi/K'] d\eta \\ &\quad - \int_{\frac{1}{2}K'}^{K'} \operatorname{cn}^2(2\eta, k') \cos[(2m+1)\eta\pi/K'] d\eta, \end{aligned}$$

i.e.,

$$b_{2m+1} = \frac{4}{K'} \int_0^{\frac{1}{2}K'} \operatorname{cn}^2(2\eta, k') \cos[(2m+1)\eta\pi/K'] d\eta. \quad (5.14)$$

On using the result due to Jacobi†

$$k^2 K^2 \operatorname{sn}^2 u = K^2 - KE - 2\pi^2 \sum_{n=1}^{\infty} \frac{nq^n}{1-q^{2n}} \cos 2nx \quad (5.15)$$

* This result is due to Jacobi. See Whittaker and Watson, "Modern Analysis," p. 510 (1920).

† Whittaker and Watson, "Modern Analysis," p. 320 (1920).

we get

$$k'^2 \operatorname{cn}^2(2\eta, k') = -k^2 + \frac{E'}{K'} + \frac{2\pi^2}{K'^2} \sum_{n=1}^{\infty} \frac{nq'^n}{1-q'^{2n}} \cos [2n\pi\eta/K']. \quad (5.16)$$

This is certainly valid for all real values of η .

On substituting for $\operatorname{cn}^2(2\eta, k')$ from (5.15) in (5.14) and integrating term by term we get

$$b_{2m+1} = \frac{4(E' - k^2 K') (-1)^m}{K' k'^2 (2m+1) \pi} + \frac{8\pi}{K'^2 k'^2} \sum_{n=1}^{\infty} \frac{nq'^n (-1)^{m+n}}{1-q'^{2n}} \cdot \frac{2m+1}{(2m+1)^2 - 4n^2}. \quad (5.17)$$

This series is very rapidly convergent so that no more than four or five terms are needed to calculate b_{2m+1} with sufficient accuracy. The values of a_{2m+1} are then easily found.

$$a_{2m+1} = (b_{2m+1} - c_{2m+1}) k'^2/k^2, \quad (5.18)$$

where

$$c_{2m+1} = \frac{\pi^2 (2m+1) q'^{(2m+1)/2}}{k'^2 K'^2 (1 - q'^{2m+1})}.$$

It follows that the function ψ is given by

$$\psi = \sum_{n=0}^{\infty} a_{2n+1} \cos [(2n+1) \pi \eta / K'] \cosh [(2n+1) \pi \xi / K'] \operatorname{sech} [(2n+1) \pi K / 2K'],$$

and consequently

$$\phi = - \sum_{n=0}^{\infty} a_{2n+1} \sin [(2n+1) \pi \eta / K'] \sinh [(2n+1) \pi \xi / K'] \operatorname{sech} [(2n+1) \pi K / 2K']. \quad (5.19)$$

5.2. The Stresses for the Case of One Indentation.

The stresses may be divided into two systems :—

$$A: \quad \widehat{xz} = -\mu\tau y, \quad \widehat{yz} = \mu\tau x.$$

$$B: \quad \widehat{xz} = \mu\tau \frac{\partial \phi}{\partial x}, \quad \widehat{yz} = \mu\tau \frac{\partial \phi}{\partial y}.$$

These may also be written

$$A: \quad \widehat{rz} = 0, \quad \widehat{\theta z} = \mu\tau r.$$

$$B: \quad \widehat{\xi z} = \mu\tau \frac{\partial \phi}{\partial \xi} \left| \frac{dt}{d\rho} \right|, \quad \widehat{\eta z} = \mu\tau \frac{\partial \phi}{\partial \eta} \left| \frac{dt}{d\rho} \right|. \quad (5.21)$$

Since the maximum shearing stress must occur on the boundary of the cross-section,* it is in the stresses on the boundary that we are chiefly interested.

The stress in the two systems must be calculated separately, but, since the resultant stress on the boundary is tangential, only the tangential components of the stresses in the two systems need be considered.

I. *Stresses due to System A.*—On the circular part of the boundary the stress is altogether tangential and of constant magnitude $\mu\tau$.

On the indentation the tangential stress is of magnitude $\mu\tau r \sin \alpha$, where α is the angle shown in fig. 1.

It follows from (4.24) that

$$\sin \alpha = \operatorname{cn} (2\eta, k') = k(1 - r^2)/2k'r,$$

so that the tangential stress is

$$\mu\tau \operatorname{cn} (2\eta, k') = \mu\tau k(1 - r^2)/2k'. \quad (5.22)$$

II. *Stresses due to System B.*—Before calculating the stress due to system B we must find the values of $\left| \frac{dt}{d\rho} \right|$ on the boundary.

It is found that on the indentation

$$\left| \frac{dt}{d\rho} \right| = \frac{1}{2k'r}, \quad (5.23)$$

and that on the circular part of the boundary

$$\begin{aligned} \left| \frac{dt}{d\rho} \right| &= \frac{1 + k \operatorname{sn}^2 \xi}{2(1 + k)(1 - k \operatorname{sn}^2 \xi)} = (\operatorname{dn} 2\xi - k \operatorname{cn} 2\xi)/2k'^2 \\ &= 1/2(1 + k^2 - 2k \cos \theta)^{\frac{1}{2}}. \end{aligned} \quad (5.24)$$

5.3. *Stresses at Particular Points.*

I. At the point P in fig. 1 the stress in system A is

$$\widehat{z\theta} = \mu\tau(1 - k')/k; \quad (5.31)$$

and in system B it is

$$\widehat{z\eta} = -\mu\tau \frac{(1 + k')\pi}{2kk'K'} \sum_{m=0}^{\infty} a_{2m+1} (2m + 1) \tanh [(2m + 1)\pi K/2K']. \quad (5.32)$$

II. At the point Q in fig. 1 the stress in system A is

$$\widehat{z\theta} = \mu\tau,$$

* Filon, *loc. cit.*, p. 340.

and in system B it is

$$\widehat{z\xi} = -\frac{\mu\tau\pi}{2K'(1-k)} \sum_{m=0}^{\infty} a_{2m+1} (-1)^m (2m+1) \operatorname{sech} [(2m+1)\pi K/2K']. \quad (5.34)$$

5.4. The Torsional Rigidity for the Case of One Indentation.

The torsional rigidity C is given by*

$$\begin{aligned} C &= \mu \iint \left(x^2 + y^2 + x \frac{\partial \phi}{\partial y} - y \frac{\partial \phi}{\partial x} \right) dx dy \\ &= \mu I + C', \end{aligned} \quad (5.41)$$

where I is the second moment of the cross-section about the axis of z , and

$$\begin{aligned} C' &= \mu \iint \left(x \frac{\partial \phi}{\partial y} - y \frac{\partial \phi}{\partial x} \right) dx dy \\ &= -\mu \int \phi \frac{\partial \phi}{\partial v} ds, \end{aligned}$$

taken over the boundary of the cross-section and in which $\partial \phi / \partial v$ denotes the derivative in the direction of the outward normal. When

$$\eta = \pm \frac{1}{2} K', \quad \frac{\partial \phi}{\partial v} = 0,$$

and when

$$\begin{aligned} \xi &= \pm \frac{1}{2} K \\ \phi \frac{\partial \phi}{\partial v} &= \left\{ \sum_{m=0}^{\infty} a_{2m+1} \sin [(2m+1)\pi \eta/K'] \tanh [(2m+1)\pi K/2K'] \right\} \\ &\quad \times \left\{ \sum_{m=0}^{\infty} a_{2m+1} \sin [(2m+1)\pi \eta/K'] (2m+1)\pi/K' \right\} \left| \frac{dt}{d\rho} \right|, \end{aligned}$$

and so

$$C' = -\frac{1}{2} \mu \pi \sum_{m=0}^{\infty} a_{2m+1}^2 (2m+1) \tanh [(2m+1)\pi K/2K']. \quad (5.42)$$

Now

$$I = \frac{1}{2} \pi - 2 \int_{r_0}^1 r^3 \theta dr = \frac{1}{2} \pi - I',$$

where

$$\cos \theta = k(1+r^2)/2r \quad \text{and} \quad r_0 = k/(1+k').$$

We find that

$$I' = \frac{1}{2} \sin^{-1} k' - \frac{k'}{2k} - \frac{3k'^3}{2k^3} + \left[\frac{k'^2}{k^2} + \frac{3k'^4}{2k^4} \right] \cos^{-1} k'. \quad (5.43)$$

* Love, *loc. cit.*, p. 312.

Having found C' and I we can get C , the torsional rigidity. It is useful to compare the torsional rigidity with that of a circular section of the same area,* and so the area of the indentation must be found. This area is

$$2 \int_{r_0}^1 r \theta dr = \sin^{-1} k' - \frac{k'}{k} + \frac{k'^2}{k^2} \cos^{-1} k'.$$

5.5. Numerical Results for a Section with One Indentation.

The numerical values of the stresses at the points P and Q and the torsional rigidity have been worked out for the case $k = \sin 70^\circ$. This gives an indentation subtending an angle of 40° at the origin and having a depth

$$1 - (1 - k')/k = 0.29979.$$

See fig. 1.

The coefficients b_{2m+1} of (5.17) are given in Table I.

Table I.

$2m+1.$	$b_{2m+1}.$	$2m+1.$	$b_{2m+1}.$	$2m+1.$	$b_{2m+1}.$	$2m+1.$	$b_{2m+1}.$
1	0.838234	15	0.000719	29	-0.000098	43	0.000030
3	0.177224	17	-0.000498	31	0.000081	45	-0.000026
5	-0.020653	19	0.000352	33	-0.000066	47	0.000023
7	0.007444	21	-0.000260	35	0.000056	49	-0.000020
9	-0.003418	23	0.000198	37	-0.000047	51	0.000018
11	0.001846	25	-0.000154	39	0.000040	53	-0.000016
13	-0.001110	27	0.000122	41	-0.000035		

The coefficients c_{2m+1} of (5.17) are found to be

$$c_1 = 2.857591, \quad c_3 = 0.066155, \quad c_5 = 0.000857, \quad c_7 = 0.000010$$

and the rest are negligible.

From (5.17) the values of the coefficients a_{2m+1} are then obtained, and on substituting these values in the equations of 5.3 the numerical results given in Table II are obtained.

Stresses.

Table II.

Point.	System A.	System B.	Total stress.
P	0.7002 $\mu\tau$	1.0175 $\mu\tau$	1.718 $\mu\tau$
Q	$\mu\tau$	-0.0236 $\mu\tau$	0.976 $\mu\tau$

* Filon, *loc. cit.*, p. 329.

Torsional Rigidity.—On substituting in 5.4 we get the value of the torsional rigidity.

$$I = \frac{1}{2}\pi - 0.1142 = 1.4566, \quad C' = -\mu 0.1118, \\ \text{and so} \quad C = \mu 1.3447. \quad (5.53)$$

The area A of the section $= 2.9946$; and if C_0 denote the torsional rigidity of a circular section of the same area,

$$C_0 = A^2/2\pi = 1.4273; \quad C/C_0 = 0.9421. \quad (5.54)$$

We may assume by comparison with Filon's results that the point P is the point of maximum stress.

If S denote the maximum stress and S_0 the maximum stress in a circular section of equal area subjected to the same twist,

$$S = \mu\tau 1.718, \quad S_0 = \mu\tau 0.9763 \quad \text{and} \quad S/S_0 = 1.760. \quad (5.55)$$

Hence the efficiency* equals

$$E = \frac{C/C_0}{S/S_0} = 0.5353. \quad (5.56)$$

We notice that the stress at Q in our section and the maximum stress in a circular section of equal area are almost exactly equal.

6. Two Indentations.

Now consider the case of two indentations. The transformation is given in (4.6) and the boundary conditions in (4.81) and (4.82).

We have†

$$\frac{k'}{k} \operatorname{cn}(2\eta, k') = \frac{2\pi}{K'k} \sum_{m=0}^{\infty} \frac{\cos[(2m+1)\eta\pi/K']}{q'^{-(2m+1)/2} + q'^{(2m+1)/2}} \quad (6.11)$$

and

$$\frac{1}{k} \operatorname{dn}(2\eta, k') - 1 = \frac{\pi}{2K'k} - 1 + \frac{2\pi}{K'k} \sum_{m=1}^{\infty} \frac{\cos(2m\eta\pi/K')}{q'^{-m} + q'^m}. \quad (6.12)$$

These are valid in particular when η is real. The expansion (6.11) is of the form required but (6.12) is not. Treating (6.12) in the same way as we treated (5.15) we obtain the series valid for $-\frac{1}{2}K' < \eta < \frac{1}{2}K'$

$$\frac{1}{k} \operatorname{dn}(2\eta, k') - 1 = \sum_{m=0}^{\infty} b_{2m+1} \cos[(2m+1)\eta\pi/K'] \quad (6.13)$$

* Filon, *loc. cit.*, p. 343.

† Whittaker and Watson, "Modern Analysis," p. 511 (1920).

where

$$b_{2m+1} = \frac{4}{\pi} \left[\frac{\pi}{2K'k} - 1 \right] \frac{(-1)^m}{2m+1} + \frac{8}{K'k} \sum_{n=1}^{\infty} \frac{(-1)^{m+n}}{q'^{-n} + q'^n} \cdot \frac{2m+1}{(2m+1)^2 - 4n^2}. \quad (6.14)$$

Then we have

$$\psi = \sum_{m=0}^{\infty} a_{2m+1} \cos [(2m+1) \pi \eta / K'] \cosh [(2m+1) \pi \xi / K'] \\ \operatorname{sech} [(2m+1) \pi K / 2K'], \quad (6.15)$$

where

$$a_{2m+1} = \frac{1}{2} (b_{2m+1} - c_{2m+1}),$$

and

$$c_{2m+1} = \frac{2\pi}{K'k} \left[q'^{-(2m+1)/2} + q'^{(2m+1)/2} \right].$$

In the same way as before we obtain the results

$$\sin \alpha = \operatorname{cn} (2\eta, k') = k(1 - r^4)/2k'r^2,$$

so that the tangential stress on the indentation due to system A is

$$\mu \tau r \operatorname{cn} (2\eta, k') = \mu \tau k (1 - r^4)/2k'r. \quad (6.21)$$

On the indentations we have

$$\left| \frac{dt}{d\rho} \right| = \frac{1}{k'r}$$

and on the circular part of the boundary

$$\left| \frac{dt}{d\rho} \right| = \frac{1}{1+k} \frac{1 + k \operatorname{sn}^2 \xi}{1 - k \operatorname{sn}^2 \xi} = \frac{1}{\sqrt{(1+k^2 - 2k \cos 2\theta)}}.$$

As before we require the area of the indentations and their second moment about the axis of z .

In this case the finding of the area involves the evaluation of an elliptic integral and it is found more convenient to use a method of numerical integration.

The second moment of the area of one indentation is found to be

$$I' = \frac{1}{4} \left[\cos^{-1} k - \frac{k'}{k} + \frac{k'^2}{k^2} \cos^{-1} k' \right].$$

The Torsional Rigidity.—The torsional rigidity is

$$C = \frac{1}{2} \mu \pi - 2 \mu I' + C', \quad (6.22)$$

where

$$C' = -\mu\pi \sum_{m=0}^{\infty} a_{2m+1}^2 (2m+1) \tanh [(2m+1) \pi K/2K'].$$

The Stress at the Point P.—In system A

$$\widehat{z\theta} = \mu\tau [(1-k')/k]^{\frac{1}{2}}. \quad (6.23)$$

In system B

$$\widehat{z\eta} = -\mu\tau \frac{\pi(1+k')^{\frac{1}{2}}}{k^{\frac{1}{2}}k'K'} \sum_{m=0}^{\infty} a_{2m+1} (2m+1) \tanh [(2m+1) \pi K/2K']. \quad (6.24)$$

Numerical Results for Two Indentations.—The case worked out is for $k = \sin 50^\circ$. See fig. 2.

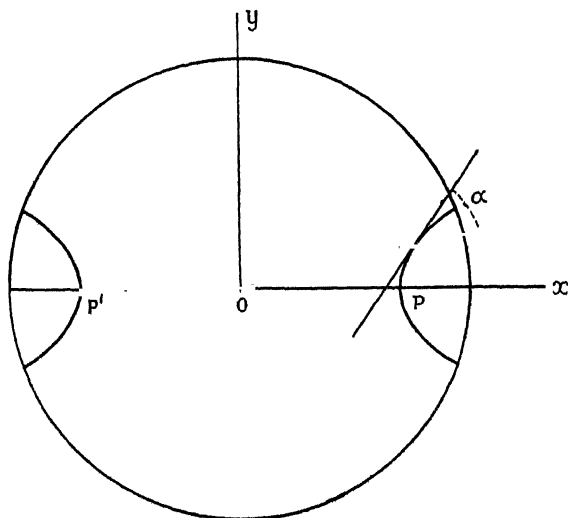


FIG. 2.

This gives two indentations each subtending an angle of 40° at the centre and having a depth $1 - r_0 = 1 - \sqrt{[(1-k')/k]} = 0.317133$. These indentations each subtend the same angle as the single indentation in the case already considered. They are, however, slightly deeper, but for a rough comparison may be taken to be of the same size and shape. The stresses at the points P and P' are in this case the maximum stresses.

The Stress at P.

Table III.

System A.	System B.	Total stress S.
$0.6829 \mu\tau$	$0.9546 \mu\tau$	$1.6375 \mu\tau$

(6.31)

By evaluating $\int_{r_0}^1 r \theta \, dr$ by Simpson's Rule, using nine ordinates, it is found that the area of each indentation is 0.1515. This is slightly greater than the area of the indentation in the first case (0.1470). The maximum stress in a circular section of the same area is

$$S = \mu \tau 0.9421 \quad \text{and so} \quad S/S_0 = 1.738. \quad (6.32)$$

The Torsional Rigidity.—We find that

$$I' = 0.1184, \quad C' = -\mu 0.2312, \quad C = \mu 1.1029, \quad (6.33)$$

$$C_0 = \mu 1.2914, \quad C/C_0 = 0.8540, \quad (6.34)$$

and so

$$E = \frac{C/C_0}{S/S_0} = 0.4908. \quad (6.35)$$

7. Comparisons of the Results for One and Two Indentations.

	S/S_0 .	C/C_0 .	E.	R/μ .
One indentation	1.760	0.9421	0.5353	0.2260
Two indentations ..	1.738	0.8540	0.4908	0.4680

where R is the reduction in torsional rigidity.

We see that the ratio of the maximum stress to the maximum stress in a circular section of equal area is approximately the same in the two cases, but that the reduction in torsional rigidity in the case of two indentations is slightly more than twice its value in the case of one indentation. The slight excess may be due to the slightly greater depth and area of the indentation.

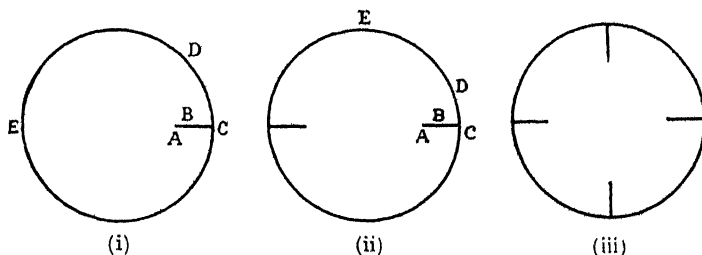


FIG. 3.

8. Circular Shafts with Evenly Placed Slits.

We now consider solutions of the torsion problem for a circular shaft with one, two or four equal slits from the surface towards the axis. The transformations and the boundary conditions have been given in 4. See fig. 3.

8.1. One Slit.

The function ψ is given by

$$\psi = \sum_{m=0}^{\infty} a_{2m+1} \cos [(2m+1) \pi \eta / K'] \cosh [(2m+1) \pi \xi / K'] \operatorname{sech} [(2m+1) \pi K / K'], \quad (8.11)$$

where

$$\frac{1}{2} [k^2 \operatorname{nd}^4 (\eta, k') - 1] = \sum_{m=0}^{\infty} a_{2m+1} \cos [(2m+1) \pi \eta / K']$$

in the range

$$-\frac{1}{2} K' < \eta < \frac{1}{2} K'.$$

It is found that

$$\operatorname{dn}^4 u = -\frac{1}{\pi} \frac{d^2}{du^2} (\operatorname{dn}^2 u) - \frac{1}{3} k'^2 + \frac{2}{3} (1 + k'^2) \operatorname{dn}^2 u. \quad (8.12)$$

From 5.15 we obtain

$$\operatorname{dn}^2 u = \frac{E}{K} + \frac{2\pi^2}{K^2} + \frac{2\pi^2}{K^2} \sum_{n=1}^{\infty} \frac{n q^n \cos 2nx}{1 - q^{2n}}$$

where $u = 2Kx/\pi$, and on substituting in 8.12 and putting $u + K$ for u we obtain

$$\frac{k'^4}{\operatorname{dn}^4 u} = \frac{2(1 + k'^2) E}{3K} - \frac{1}{3} k'^2 + \sum_{n=1}^{\infty} \left[\frac{4\pi^2}{3K^2} (1 + k'^2) n + \frac{\pi^4 n^3}{3K^4} \right] \frac{(-q)^n}{1 - q^{2n}} \cos 2nx$$

and so

$$\begin{aligned} \frac{1}{2} \left[\frac{k^2}{\operatorname{dn}^4 (\eta, k')} - 1 \right] &= \left[\frac{(1 + k^2) E'}{3k^2 K'} - \frac{2}{3} \right] \\ &+ \sum_{n=1}^{\infty} \frac{4\pi^2}{3K'^2 k^2} \left[\frac{\pi^2 n^3}{K'^2} + (1 + k^2) n \right] \frac{\cos (2n\pi \eta / K')}{q'^{-2n} - q'^{2n}} \\ &- \sum_{n=0}^{\infty} \left[\frac{\pi^2}{6K' k^2} \left[\frac{(2n+1)^3 \pi^2}{K'^2} + 4(1 + k^2)(2n+1) \right] \frac{\cos [(2n+1) \pi \eta / K']}{q'^{-2n-1} - q'^{2n+1}} \right]. \end{aligned} \quad (8.13)$$

These expansions are certainly valid for all real values of η , but only the second is of the required form. Treating the first series in the same way as before, we get a series valid when $-\frac{1}{2} k' < \eta < \frac{1}{2} K'$ and finally we obtain the following result

$$\frac{1}{2} [k^2 \operatorname{nd}^4 (\eta, k') - 1] = \sum_{m=0}^{\infty} (b_{2m+1} - c_{2m+1}) \cos [(2m+1) \pi \eta / K'], \quad (8.14)$$

where

$$b_{2m+1} = \frac{4}{\pi} \left[\frac{(1+k^2)E'}{3k^2K'} - \frac{2}{3} \right] \frac{(-1)^m}{2m+1} \\ + \frac{16\pi}{3K'^2k^2} \sum_{n=1}^{\infty} \left[\frac{n^3\pi^2}{K'^2} + (1+k^2)n \right] \frac{(-1)^{m+n}}{q'^{-2n} - q'^{2n}} \cdot \frac{2m+1}{(2m+1)^2 - 4n^2}$$

and

$$c_{2m+1} = \frac{\pi^2}{6K'^2} \left[\frac{(2m+1)^3\pi^2}{K'^2} + 4(1+k^2)(2m+1) \right] \frac{1}{q'^{-2m-1} - q'^{2m+1}}.$$

The series for b_{2m+1} is similar to those found in the two previous cases and is very rapidly convergent.

By a method similar to that already used we find that the reduction in torsional rigidity is

$$\frac{1}{2}\mu\tau \sum_{m=0}^{\infty} a_{2m+1}^2 (2m+1) \tanh [(2m+1)\pi K/K']. \quad (8.15)$$

8.2. Two Slits.

The function ψ is given by (8.11), where in this case

$$\frac{1}{2}[k \operatorname{nd}^2(\eta, k') - 1] = \sum_{m=0}^{\infty} a_{2m+1} \cos [(2m+1)\pi\eta/K']$$

in the range

$$-\frac{1}{2}K' < \eta < \frac{1}{2}K'.$$

It is found that

$$a_{2m+1} = b_{2m+1} - c_{2m+1},$$

where

$$b_{2m+1} = \frac{2}{\pi} \left[\frac{E'}{K'k} - 1 \right] \frac{(-1)^m}{(2m+1)} + \frac{4\pi}{K'^2k} \sum_{n=1}^{\infty} \frac{2n(-1)^{m+n}}{q'^{-2n} - q'^{2n}} \cdot \frac{2m+1}{(2m+1)^2 - 4n^2} \quad (8.21)$$

and

$$c_{2m+1} = \frac{\pi^2}{K'^2k} \frac{2m+1}{q'^{-2m-1} - q'^{2m+1}}. \quad (8.22)$$

In this case the reduction in torsional rigidity is

$$\mu\tau \sum_{m=0}^{\infty} a_{2m+1}^2 (2m+1) \tanh [(2m+1)\pi K/K']. \quad (8.23)$$

8.3. Four Slits.

The function ψ is given by (8.11) where

$$\frac{1}{2}[k^4 \operatorname{nd}(\eta, k') - 1] = \sum_{m=0}^{\infty} a_{2m+1} \cos [(2m+1)\pi\eta/K']$$

in the range

$$-\frac{1}{2}K' < \eta < \frac{1}{2}K'.$$

It is found that

$$a_{2m+1} = b_{2m+1} - c_{2m+1},$$

where

$$b_{2m+1} = \left[\frac{1}{k'k^{\frac{1}{2}}} - \frac{2}{\pi} \right] \frac{(-1)^m}{(2m+1)} + \frac{4}{k'k^{\frac{1}{2}}} \sum_{n=1}^{\infty} \frac{(-1)^{m+n}}{q'^{-2n} + q'^{2n}} \cdot \frac{2m+1}{(2m+1)^2 - 4n^2} \quad (8.31)$$

and

$$c_{2m+1} = \frac{\pi}{K'k^{\frac{1}{2}}} \frac{1}{q'^{-2m-1} + q'^{2m+1}} \quad (8.32)$$

In this case the reduction in torsional rigidity is

$$2\mu\pi \sum_{m=0}^{\infty} a_{2m+1}^2 (2m+1) \tanh [(2m+1) \pi K/K']. \quad (8.33)$$

9. Values of the Torsional Rigidity for One, Two and Four Slits.

Let d be the depth of the slit, and T the percentage reduction in torsional rigidity.

Table IV.

	k .	d .	R/μ .	T .
One slit	$\sin 45^\circ$	0.2929	0.1088	6.928
One slit	$\sin 30^\circ$	0.5	0.2694	17.148
Two slits	$\sin 45^\circ$	0.1591	0.0707	4.502
Two slits	$\sin 30^\circ$	0.2929	0.2146	13.659
Four slits	$\sin 30^\circ$	0.1591	0.1394	8.873

10. Comparisons of the Results for the Torsional Rigidity.

From Table IV we see that the effect of two slits is slightly less than twice that of one slit and that the effect of four slits is slightly less than twice the effect of two slits. We may then assume that the reduction of torsional rigidity due to one slit of depth 0.1591 is slightly more than half 4.502 per cent., *i.e.*, about 2.3 per cent. If we take Filon's result* for a slit reaching the axis we can construct a graph from the following data.

d	0.16	0.29	0.5	1.0
T	2.3	6.9	17	44

* Filon, *loc. cit.*, p. 350. In this limiting case his ellipse becomes a circle.

The other two results of Filon's are also marked on the graph, fig. 4, showing close agreement. This verifies the soundness of his assumption that the results for a circle could be inferred from those of an ellipse.

On comparing the reduction in torsional rigidity for one indentation with that for one slit of almost the same depth (the first case) we see that the former is slightly more than twice the latter. If, however, we compare the quantity $(-C')$ with the reduction in the case of the slit we see that they are nearly equal. The further reduction in the case of the indentation is due to the decrease in the second moment of the cross-section about the axis. It

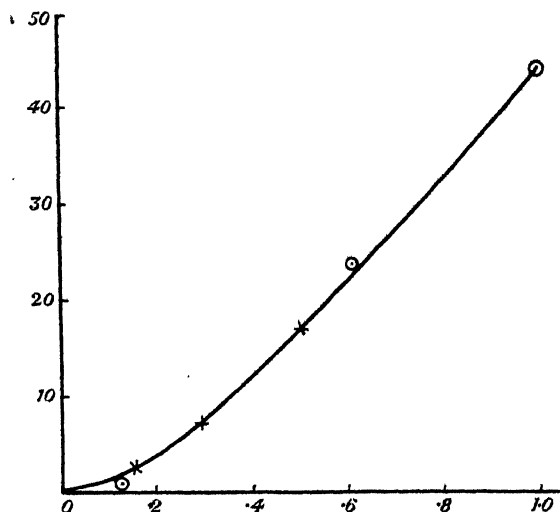


Fig. 4.—The points marked \odot are Filon's results for ellipses.

seems reasonable therefore to suppose that, for an indentation of shape something between a semicircle and a slit, the torsional rigidity may be taken to be proportional to the second moment about the axis less the reduction due to a slit of the same depth as the indentation.

11. *The Distortion of the Cross-sections in the Direction of the Axis.*

The distortion of the cross-sections at right angles to their planes is given for the shafts with slits by

$$w = \tau\phi = -\tau \sum_{m=0}^{\infty} a_{2m+1} \sin [(2m+1)\pi\eta/K'] \sinh [(2m+1)\pi\xi/K'] \operatorname{sech} [(2m+1)\pi K/K'].$$

This displacement is calculated in the case of one slit of depth 0.2929 (*i.e.*, $k = \sin 45^\circ$) at the points A, B, C, D, E of fig. 3 (i) and is given in Table V.

Table V.

Point.	A(K, 0).	B(K, K'/4).	C(K, $\frac{1}{2}K'$).	D($\frac{1}{2}K$, $\frac{1}{2}K'$).	E(0, $\frac{1}{2}K'$).
w/τ	0	0.4906	0.5615	0.0394	0

The displacement at points of the boundaries of the slit is shown in fig. 5 and at points of the circular part of the boundary in fig. 6. These graphs have been constructed from the above table and a knowledge of $\partial w/\partial s$ at these points.

The most noticeable feature is the way in which the displacement is confined to the neighbourhood of the slit. This is clearly shown in fig. 6.

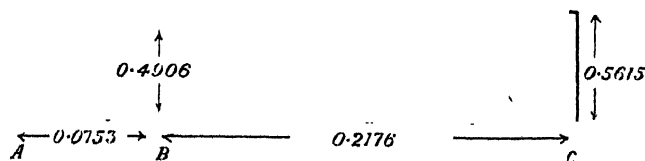


FIG. 5.



FIG. 6.

The relative shift of the two sides at the lips of the slit has been calculated in other cases giving the results shown in Table VI.

Table VI.

	k .	Depth of slits.	Shift.	Shift/depth of slits.
One slit	$\sin 45^\circ$	0.2929	$\tau 1.123$	$\tau 3.835$
One slit	$\sin 30^\circ$	0.5	$\tau 1.855$	$\tau 3.910$
Two slits	$\sin 45^\circ$	0.1591	$\tau 0.620$	$\tau 3.896$
Two slits	$\sin 30^\circ$	0.2929	$\tau 1.107$	$\tau 3.781$
Four slits	$\sin 30^\circ$	0.1591	$\tau 0.611$	$\tau 3.841$

The Table VI shows that ratio (shift/depth of slits) is almost independent of the depth or number of the slits.

12. *Flexure—Statement of the Notation.*

We take the axes of co-ordinates as in the first part of the paper and suppose that the load W is applied at the end $z = 0$ in the negative direction of the axis of y .

We consider the cases of one and two slits in the plane $y = 0$. Apart from rigid body displacements the displacements are given by*

$$u = -\frac{W}{EI} \sigma xyz - \tau yz \quad (12.11)$$

$$v = \frac{W}{EI} \left\{ \sigma \frac{z}{2} (x^2 - y^2) - \frac{z^3}{6} \right\} + \tau zx \quad (12.12)$$

$$w = \frac{W}{EI} \{ \chi + \frac{1}{2} yz^2 - x^2 y \} + \tau \phi, \quad (12.13)$$

where E is Young's Modulus and σ is Poisson's Ratio. I is the second moment of the cross-section about Ox and in all the sections treated has the value $\frac{1}{4}\pi$. (In the first part of the paper I is the second moment about Oz .)

The function χ is independent of z and satisfies the equation

$$\nabla^2 \chi = 0 \quad (12.21)$$

at all points of the interior of the cross-section and satisfies the condition

$$\frac{\partial \chi}{\partial \nu} = lxy(2 + \sigma) + m[x^2 - \frac{1}{2}\sigma(x^2 - y^2)] \quad (12.22)$$

at all points of the boundary. The outward normal is denoted by ν and $(l, m, 0)$ are its direction cosines.

The function χ must be determined to satisfy these conditions. ϕ is the torsion function and τ is the twist which must be determined from the condition that there is no couple about the axis of the cylinder.

13. *Method of Solution.*

For a circular cross-section with no slits the function χ is given by†

$$\chi = \chi_0 = \frac{1}{4}(3 + 2\sigma)y + \frac{1}{4}(3x^2y - y^3). \quad (13.1)$$

Put

$$\chi = \chi_0 + \chi_1. \quad (13.2)$$

* Love, *loc. cit.*, p. 334. The notation is a little different.

† Love, *loc. cit.*, p. 335.

Then the boundary conditions for χ_1 are

$$\frac{\partial \chi_1}{\partial \nu} = 0 \quad (13.3)$$

on the circular part of the boundary, and

$$\frac{\partial \chi_1}{\partial y} = -\frac{1}{4}(3 + 2\sigma) + \frac{1}{4}x^2(1 - 2\sigma) \quad (13.4)$$

on the slit or slits since $\partial/\partial \nu$ and m change sign together.

14. The Function χ_1 for Two Slits.

As in the case of torsion the transformation used is

$$\rho = x + iy = k^{\frac{1}{2}} \operatorname{sn}(\xi + i\eta) = k^{\frac{1}{2}} \operatorname{sn} t. \quad (14.1)$$

In this case the function χ_1 is even in ξ and odd in η and the conditions become :
when $\eta = \pm \frac{1}{2}K'$,

$$\frac{\partial \chi_1}{\partial \eta} = 0, \quad (14.2)$$

when $\xi = K$,

$$\frac{\partial \chi_1}{\partial \xi} = \pm \frac{1}{4}[(3 + 2\sigma) - (1 - 2\sigma)x^2] \left| \frac{d\rho}{dt} \right|,$$

the positive sign being taken when $\eta > 0$ and the negative sign when $\eta < 0$.
This second condition is equivalent to

$$\begin{aligned} \left(\frac{\partial \chi_1}{\partial \xi} \right)_{\xi=K} &= \frac{1}{4}[(3 + 2\sigma) - (1 - 2\sigma)k \operatorname{nd}^2(\eta, k')] k^{\frac{1}{2}} k'^2 \operatorname{sn}(\eta, k') \operatorname{cn}(\eta, k') \operatorname{nd}^2(\eta, k'), \\ \text{i.e.,} \quad &= \frac{1}{4} k^{\frac{1}{2}} \left\{ \left[(3 + 2\sigma) - \frac{(2 - k'^2)(1 - 2\sigma)}{6k} \right] \frac{d}{d\eta} [\operatorname{nd}(\eta, k')] \right. \\ &\quad \left. + \frac{(1 - 2\sigma)}{6k} \frac{d^3}{d\eta^3} [\operatorname{nd}(\eta, k')] \right\}. \end{aligned} \quad (14.3)$$

On using the Fourier expansion* for $\operatorname{nd}(\eta, k')$ we find that

$$\left(\frac{\partial \chi_1}{\partial \xi} \right)_{\xi=K} = \sum_{n=1}^{\infty} \alpha_n \sin \left[\frac{n\pi\eta}{K'} \right] \quad (14.4)$$

$$\begin{aligned} &= \frac{\pi^2}{2k^{\frac{1}{2}}K'^2} \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{q'^n + q'^{-n}} \left\{ \left[(3 + 2\sigma) - \frac{(2 - k'^2)(1 - 2\sigma)}{6k} \right] n \right. \\ &\quad \left. - \frac{(1 - 2\sigma)}{6k} \cdot \frac{\pi^2}{K'^2} \cdot n^3 \right\} \sin \left[\frac{n\pi\eta}{K'} \right]. \end{aligned} \quad (14.5)$$

* Whittaker and Watson, "Modern Analysis," p. 511 (1920).

We require the expansion in odd multiples of $(\eta\pi/K')$, so proceeding as before we have when $-\frac{1}{2}K' < \eta < \frac{1}{2}K'$

$$\sum_{n=1}^{\infty} \alpha_n \sin [n\pi\eta/K'] = \sum_{m=0}^{\infty} (\alpha_{2m+1} + c_{2m+1}) \sin [(2m+1)\pi\eta/K'],$$

where

$$c_{2m+1} = \frac{8}{\pi} \sum_{n=1}^{\infty} (-1)^{m+n} \alpha_{2n} \cdot \frac{n}{(2m+1)^2 - 4n^2}. \quad (14.6)$$

It follows that if

$$a_{2m+1} = \alpha_{2m+1} + c_{2m+1}, \quad (14.7)$$

$$\chi_1 = \frac{K'}{\pi} \sum_{m=0}^{\infty} a_{2m+1} \frac{1}{2m+1} \cosh [(2m+1)\xi\pi/K'] \sin [(2m+1)\eta\pi/K'] \operatorname{cosech} [(2m+1)\pi K/K']. \quad (14.8)$$

15. The Function χ_1 for One Slit.

The transformation is

$$\rho = x + iy = k \operatorname{sn}^2 (\xi + i\eta) = k \operatorname{sn}^2 t. \quad (15.1)$$

In this case χ_1 is odd in ξ and odd in η and the boundary conditions become:—when $\eta = \pm \frac{1}{2}K'$,

$$\frac{\partial \chi_1}{\partial \eta} = 0,$$

when $\xi = K$,

$$\begin{aligned} \frac{\partial \chi_1}{\partial \xi} &= \frac{k}{4} \left\{ (3 + 2\sigma) \frac{d}{d\eta} [\operatorname{nd}^2 (\eta, k')] - (1 - 2\sigma) \frac{k^2}{3} \frac{d}{d\eta} [\operatorname{nd}^6 (\eta, k')] \right\}, \\ \text{i.e.,} \\ &= \frac{k}{4} \left\{ \left[(3 + 2\sigma) - (1 - 2\sigma) \cdot \frac{8(2 - k'^2)^2 - 9k^2}{45k^2} \right] \frac{d}{d\eta} \right. \\ &\quad \left. + \frac{(1 - 2\sigma)(2 - k'^2)}{18k^2} \frac{d^3}{d\eta^3} - \frac{(1 - 2\sigma)}{360k^2} \frac{d^5}{d\eta^5} \right\} \operatorname{nd}^2 (\eta, k'). \end{aligned} \quad (15.2)$$

From (5.15) we deduce that

$$\operatorname{nd}^2 (\eta, k') = \frac{E'}{k^2 K'} + \frac{2\pi^2}{k^2 K'^2} \sum_{n=1}^{\infty} \frac{(-1)^n n}{q'^{-n} - q'^n} \cos [n\pi\eta/K'],$$

so that

$$\begin{aligned} \left(\frac{\partial \chi_1}{\partial \xi} \right)_{\xi=K} &= \frac{\pi^3}{2kK'^3} \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{q'^{-n} - q'^n} \left\{ \left[(3 + 2\sigma) - (1 - 2\sigma) \frac{8(2 - k'^2)^2 - 9k^2}{45k^2} \right] n^2 \right. \\ &\quad \left. - \frac{(1 - 2\sigma)(2 - k'^2)\pi^2 n^4}{18k^2 K'^2} - \frac{(1 - 2\sigma)\pi^4 n^6}{360k^2 K'^4} \right\} \sin \left[\frac{n\pi\eta}{K'} \right]. \end{aligned}$$

The series for $\partial\chi_1/\partial\xi$ is treated as in the case of two slits giving

$$\left(\frac{\partial\chi_1}{\partial\xi}\right)_{\xi=K} = \sum_{m=0}^{\infty} a_{2m+1} \sin [(2m+1)\pi\eta/K'] \quad (15.3)$$

valid in the range $-\frac{1}{2}K' < \eta < \frac{1}{2}K'$.

It follows that

$$\chi_1 = \frac{K'}{\pi} \sum_{m=0}^{\infty} a_{2m+1} \frac{1}{2m+1} \sinh [(2m+1)\xi\pi/K'] \sin [(2m+1)\eta\pi/K'] \operatorname{sech} [(2m+1)\pi K/K']. \quad (15.4)$$

16. Further Consideration of the Case of Two Slits.

From the symmetry of the problem we see that there is no twist, *i.e.*, $\tau = 0$. The coefficients of the series for χ_1 have been worked out for $k = \sin 30^\circ$ and $k = \sin 45^\circ$, taking $\sigma = 0.3$.

The cross-sections are distorted from their original plane state and are also tilted so that they no longer cross the strained central line at right angles. The tangent plane at the centre of the cross-section when strained makes an angle $\cos^{-1} s_0$ with the strained central line where*

$$s_0 = \frac{W}{EI} \left(\frac{\partial\chi}{\partial y} \right)_0 \quad (16.11)$$

$\left(\frac{\partial\chi}{\partial y} \right)_0$ denoting the value of $\partial\chi/\partial y$ at the centre of the cross-section.

$$s_0 = \frac{W}{EI} \left[k^{-1} \left(\frac{\partial\chi_1}{\partial\eta} \right)_0 + \left(\frac{\partial\chi_0}{\partial y} \right)_0 \right]. \quad (16.12)$$

The second term is that for a circular section with no slits and the first is the addition due to the slits.

We find that

$$\left(\frac{\partial\chi_1}{\partial\eta} \right)_0 = \sum_{m=0}^{\infty} a_{2m+1} \operatorname{cosech} [(2m+1)\pi K/K']$$

$$k^{-1} \left(\frac{\partial\chi_1}{\partial\eta} \right)_0 = 0.0228 \text{ when } k = \sin 45^\circ$$

and

$$= 0.0884 \text{ when } k = \sin 30^\circ$$

$$\left(\frac{\partial\chi_0}{\partial y} \right)_0 = 0.9 \text{ in both cases.}$$

The relation between s_0 and the depth of the slits is shown in Table VII and in fig. 7.

* Love, *loc. cit.*, p. 338.

Table VII.

k .	Depth of slits.	s_0 .
1	0	1.146
$\sin 45^\circ$	0.1591	1.175
$\sin 30^\circ$	0.2929	1.258

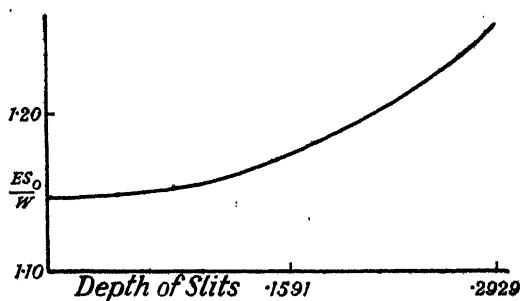


FIG. 7.

The function χ only appears in the displacement w so that u and v are not affected by the presence of the slits since τ is zero. The additional displacement w_1 due to the slits is given by

$$w_1 = \frac{W}{EI} \chi_1 = \frac{W}{E} \frac{4}{\pi} \chi_1.$$

As in the case of torsion it is found impracticable to work out the values of χ_1 for points in the interior of the cross-section, but the displacement w_1 at the points A, B, C, D, E, fig. 3 (ii) has been worked out in the case $k = \sin 45^\circ$ when the depth of the slits is 0.1591 and the results are given in Table VIII.

Table VIII.

Point.	A.	B.	C.	D.	E.
Ew_1/W	0	0.1168	0.1697	0.0362	0.0144

These values together with the values of $\partial\chi_1/\partial s$ at these points have been used in constructing the graphs. See figs. 8 and 9.

The displacements on the other parts of the boundary are obtained from the fact that χ_1 is odd in y and even in x .

As in the case of torsion we see that the disturbance due to the slits is very local, being confined to their immediate neighbourhood. The displacement w_1 on the line $x = 0$ is easily seen to be approximately of the form

$$w_1 = A \sin \frac{1}{2} (\pi y).$$

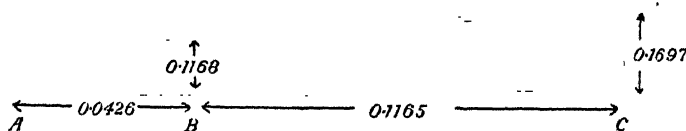


FIG. 8.

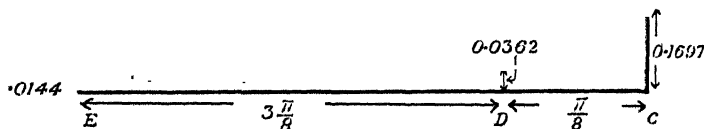


FIG. 9.

This is very similar to the displacement of the shaft without slits and merely causes a slight increase in the displacement.

The value of $\frac{Ew_1}{W}$ has also been worked out for the point C, fig. 3 (ii), in the case $k = \sin 30^\circ$, and we find that $\frac{Ew}{W} = 0.3479$.

The relative shift at the lips of the slits will be given by twice these values.

Table IX.

Relative shift.	Depth of slits.
0	0
$\frac{W}{E} 0.3394$	0.1591
$\frac{W}{E} 0.6958$	0.2929

These values, see fig. 10, show that the shift is very roughly proportional to the depth of the slits for slits up to about $0.3 \times$ radius.

17. *Further Consideration of the Case of One Slit.*

In this case there is no symmetry about the plane $x = 0$ so that a twist is involved.

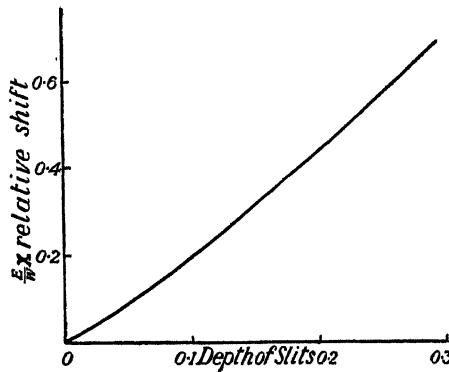


FIG. 10.

We must find τ so that*

$$\tau C + \frac{W\mu}{EI} \iint \left(y \frac{\partial \chi_1}{\partial x} - x \frac{\partial \chi_1}{\partial y} \right) dx dy = 0, \quad (17.11)$$

where C is the torsional rigidity of the shaft.

Denote the integral by J , then

$$J = \iint \left(y \frac{\partial \chi_1}{\partial x} - x \frac{\partial \chi_1}{\partial y} \right) dx dy = \int \chi_1 [y \cos(x, \nu) - x \cos(y, \nu)] ds. \quad (17.12)$$

This integral vanishes over the circular part of the boundary so that

$$J = 2 \int_k^1 (\chi_1)_{\nu=0} x dx = \int_{x=k}^1 (\chi_1)_{\nu=0} d(x^2).$$

We have

$$x^2 = k^2 n d^4(\eta, k') = 1 + 2 \sum_{m=0}^{\infty} a'_{2m+1} \cos[(2m+1)\pi\eta/K']$$

where a'_{2m+1} refers to the coefficients of (8.11).

We have also when $y = 0$ and $k < x < 1$

$$\chi_1 = \frac{K'}{\pi} \sum_{m=0}^{\infty} \frac{a_{m+1}}{2m+1} \tanh[(2m+1)\pi K/K'] \sin[(2m+1)\pi\eta/K'].$$

* Love, *loc. cit.*, p. 332.

On substituting these in the integral we get

$$J = -\frac{K'}{2} \sum_{m=0}^{\infty} a_{2m+1} a'_{2m+1} \tanh [(2m+1)\pi K/K'],$$

when $k = \sin 45^\circ$, $J = 0.2039$ and when $k = \sin 30^\circ$, $J = 0.4634$.

The values of C for these two cases are respectively $\mu 1.462$ and $\mu 1.304$. Putting these values in the equation

$$\tau = -\frac{W\mu J}{EIC}$$

we find that when $k = \sin 45^\circ$, $\tau = -\frac{W}{E} 0.1176$, and when $k = \sin 30^\circ$

$$\tau = -\frac{W}{E} 0.4534.$$

The twist is such that the side of the shaft into which the slit is cut is turned in the direction of the bending force.

The additional displacement, in the direction of Oz , due to the slit is

$$w_1 = w_2 + w_3$$

where

$$w_2 = \frac{W}{EI} \chi_1,$$

and

$$w_3 = \tau \phi.$$

In the case $k = \sin 45^\circ$ when the depth of the slit is 0.2929 this displacement has been calculated for the points A, B, C, D, E, fig. 3 (i). They are shown in Table X and by graphs in figs. 11 and 12.

Table X.

	A.	B.	C.	D.	E.
$\frac{E}{W} w_2$	0	0.2143	0.3242	0.0625	0
$\frac{E}{W} w_3$	0	-0.0871	-0.0997	-0.0070	0
$\frac{E}{W} w_1$	0	0.1272	0.2245	0.0555	0

The displacement at the point C when $k = \sin 30^\circ$ has also been calculated and we find that

$$\frac{E}{W} w_2 = 0.4871 \quad \text{and} \quad \frac{E}{W} w_3 = -0.4205,$$

so that

$$\frac{E}{W} w_1 = 0.0666.$$

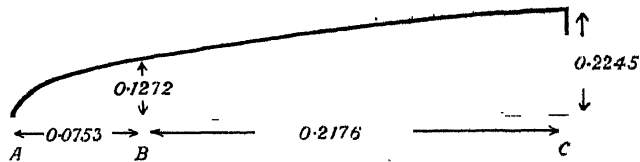


FIG. 11.

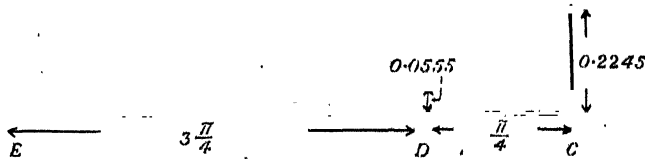


FIG. 12.

We see that the displacement due to the twist is very much increased as compared with the first case so that the relative shift at the lips of the slit is less than in the case of the slit of less depth. On the face of it this may appear strange, but is only to be expected when we consider the decreased resistance to torsion of the more deeply cut shaft.

We have considered only cases in which the bending force is at right angles to the plane of the slits, but the results for force in any transverse direction can be easily deduced. If the slits had been in the plane yOz instead of the plane zOx the solution would have been given by the function χ_0 since this gives no stress across yOz . The bending force must then be resolved parallel to Ox and Oy and two solutions added together.

The Oxidation of Sulphur at Low Pressures.

By A. RITCHIE, B.Sc., Ph.D., and E. B. LUDLAM, M.A., D.Sc., Carnegie Teaching Fellow, Edinburgh University.

(Communicated by J. Kendall, F.R.S.—Received August 2, 1932.)

The reaction between sulphur vapour and oxygen has been examined by Norrish and Rideal* for pressures of oxygen from 0.1 to 1 atmosphere at temperatures from 200° to 400° C. Between the temperatures of 235° and 305° C., they found that the reaction was confined to the sulphur surface and the walls of the vessel, while below 200° C. there was no appreciable reaction. Several investigators† have found that the oxidation of sulphur is accompanied by phosphorescence. Emeléus‡ studied the phosphorescent combustion in order to examine the reaction products, which he found consisted of sulphur dioxide and a small amount of the trioxide. During the course of the investigation he found that *sulphur dioxide* and certain organic vapours inhibit the glow, and he concluded that the oxidation is a chain reaction similar to the phosphorescent oxidation of phosphorus. Semenoff and Rjabini§ have investigated the reaction between sulphur vapour and oxygen at fairly low pressures of oxygen and have obtained very interesting results. Their reaction vessel, which contained the solid sulphur, was maintained at a temperature between 80° and 120° C. and oxygen was introduced to a pressure of 20 mm. of mercury, or less. Any SO₂ or SO₃ formed was immediately condensed in a side tube which dipped into liquid air. Under these conditions they observed in general *no reaction* between sulphur vapour and oxygen. If, however, an electric discharge were passed through the vessel a reaction started which continued *after the removal* of the discharge. This reaction was accompanied by a luminescence, and its rate was practically independent of the pressure of oxygen when this was below 1 mm., until an oxygen pressure of about 0.2 mm. was reached, when the reaction ceased. It was found that the residual oxygen pressure varied with the temperature of the reaction vessel, but was not changed by addition of inert gas or by changing the diameter of the vessel as is the case with the residual oxygen pressure in the oxidation of phosphorus.

* 'J. Chem. Soc.,' vol. 123, p. 3202 (1923).

† K. Heumann, 'Ber. deuts. chem. Ges.,' vol. 16, p. 139 (1888), and Watson, 'Chem. News,' vol. 108, p. 787 (1913).

‡ 'J. Chem. Soc.,' p. 1942 (1928).

§ 'Z. Phys. Chem.,' B, vol. 1, p. 122 (1928).

As the investigation* of the oxidation of phosphorus below the lower critical limit had proved interesting, it was decided to carry out a further examination of the sulphur oxidation.

Experimental.

To find if it were possible to obtain a stable chain oxidation of the sulphur vapour, the apparatus shown in fig. 1 was used. R is the reaction vessel, which is divided by the ground joint J, so that various filaments can be easily placed in the vessel. Borated copper wire passes through a pinch seal in the

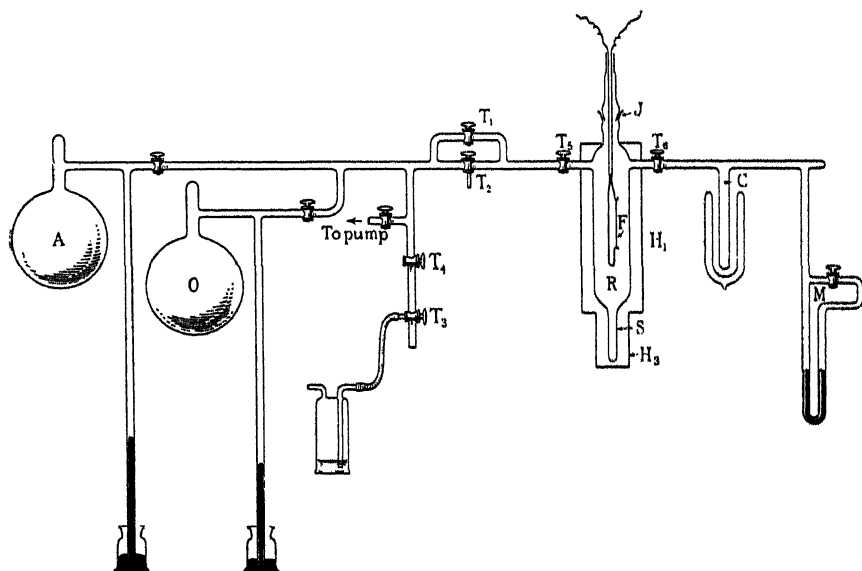


FIG. 1.—Diagram of apparatus—first type of reaction vessel.

upper part of joint J, and is soldered to degassed nickel wire, which serves as a support for the various filaments. The vessel is enclosed in an electric oven H_1 , so that the temperature can be maintained at about 120°C . The small tube S contains solid sulphur, and is enclosed in a small electric oven H_3 . M is a sulphuric acid manometer which is read by means of a microscope with an eyepiece scale (1 division of the eyepiece scale corresponds to 0.024 mm. Hg). T_2 is a small "tap-pipette" to allow small quantities of oxygen to be transferred from the oxygen reservoir O into the reaction vessel. A is a reservoir for gases other than oxygen. C is a side tube which is placed in liquid air, and thus serves as a condenser for any SO_2 and SO_3 formed.

* 'Proc. Roy. Soc.,' A, vol. 135, p. 315 (1932).

The apparatus is evacuated by means of a mercury condensation pump backed by a hyvac oil pump, and the sulphur is purified by distilling it over to the top part of the vessel from tube S, and then back to S again by use of the heaters H_1 and H_3 . The reservoir O is then filled with dry oxygen by means of taps T_3 and T_4 . Experiments are carried out by introducing a small quantity of oxygen into the reaction vessel by pipette T_2 , tap T_1 being kept closed. C is placed in liquid air and H_1 and H_3 kept at 120°C . and 90°C . respectively. The filament is now heated for a certain time and the change of pressure noted. The temperature of the filament is kept as constant as possible in a series of experiments by placing it in a Wheatstone bridge, so that the current can be regulated to give the filament a constant resistance. So as to be certain that any change of pressure could not be due to compound formation with the material of the filament, one whose surface is entirely of pyrex glass is used. The method of making this "pyrex" filament is described in a paper by Ritchie, Brown and Muir.*

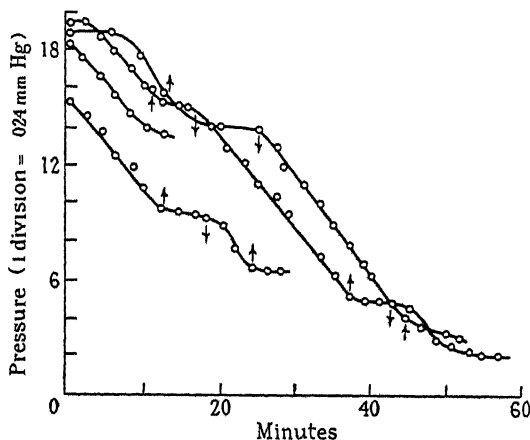


FIG. 2.—Rate of reaction with hot filament. The arrows show when the filament is switched on \downarrow and off \uparrow .

Using oxygen pressures of about 0.4 mm. of mercury it was found that a reaction took place as long as the filament was kept hot. As soon as the filament was switched off the reaction ceased. Also, the reaction apparently ceased when the oxygen pressure fell to about 0.05 mm. The results of typical series of experiments are shown in fig. 2. The arrows show the time when filament was switched on (\downarrow) and off (\uparrow). To find if the same reaction would take place from a hot silica surface, the reaction vessel shown in fig. 1

* 'Proc. Roy. Soc.,' vol. 137, p. 511 (1932).

was replaced by that shown in fig. 3. The vessel consists of a silica tube 19 cm. in length and 4 cm. in diameter. Through this tube there passes an axial tube 1 cm. in diameter. The volume of the vessel is approximately 50 c.c. The inner tube contains a coil of wire H_1 which can be heated electrically to 500°C . The temperature of the wall of the inner tube is read by means of a "platinum-rhodium" thermocouple. The main part of the reaction vessel is in an electric heater H_2 , so that the outer walls of the vessel can be kept at 200°C . The small side tube S contains the solid sulphur, and is heated to $80^\circ\text{--}100^\circ\text{C}$. by the electric heater H_3 . The vessel is connected to the rest of the apparatus by a silica to glass ground joint. The experiment was carried out in the same way as before except that, during the heating of H_1 , tap T_6 was kept closed, so that SO_2 was not removed from the vessel. The experiment showed that reaction took place only when the inner tube was raised to a temperature of about 500°C ., no reaction taking place when the walls were kept at 200°C . The rate of the reaction was slightly increased when the temperature of the axial tube was raised to 600°C . Argon appeared to have no effect on the rate of the reaction. If the product SO_2 should act as a chain-breaker it would be expected that the length of the reaction chains would be short when the experiment was carried out in the above manner. Proof that the product did, in fact, act as a chain breaker, was shown by experiments carried out with a third type of reaction vessel.

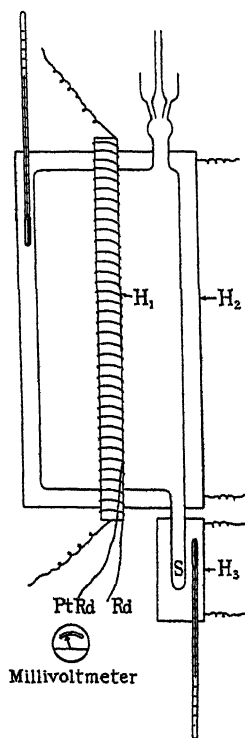


FIG. 3. — Silica reaction vessel showing the axial tube which can be heated to 500°C .

This apparatus, which is similar to the last, is made of pyrex and is shown diagrammatically in fig. 4. The vessel is 29 cm. in length and 6 cm. in diameter, the axial tube being 3 mm. in external diameter. The pyrex vessel is first wound with sheet copper and then with asbestos. Resistance ribbon is now wound round the asbestos, and this is followed by another layer of asbestos. The walls of the vessel could thus be heated to $100^\circ\text{--}150^\circ\text{C}$., the temperature being determined by a copper constantan thermocouple. A length of resistance ribbon passes through the inner tube and can be heated to 500°C ., the temperature being determined

by means of a "platinum-rhodium" thermocouple. To facilitate the removal of SO_2 and SO_3 from the reaction vessel tap T_6 was removed and the side tube C was made larger and placed as near as possible to R and, to obtain an accurate value for low oxygen pressures, a McLeod gauge was added to the apparatus. Under these conditions it was found that it was only necessary to heat the

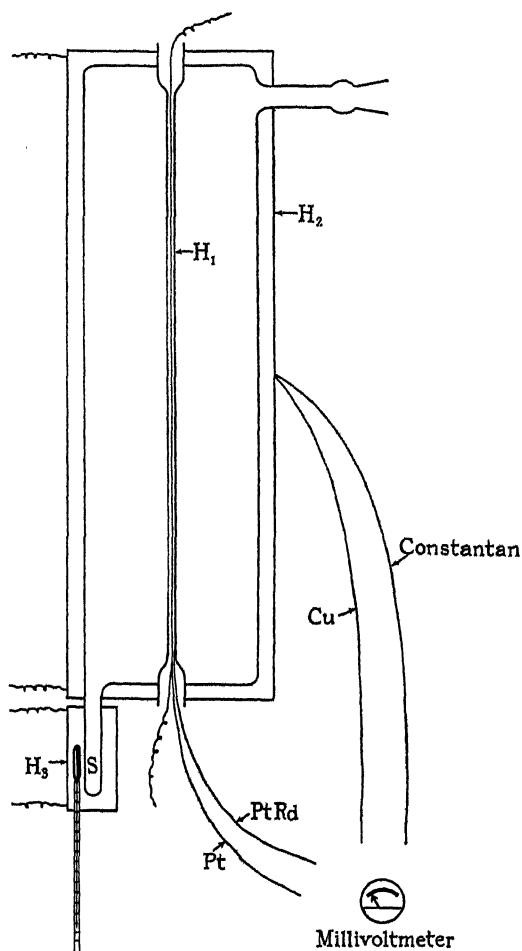


FIG. 4.—Pyrex reaction vessel showing the axial tube which is heated to 450°C . to start the reaction.

axial tube for a few minutes to start a reaction which proceeded until the pressure of oxygen fell to about 0.05 mm . In looking through a window at the top of the vessel a blue luminescence was observed during the reaction. Fig. 5 shows the course of the reaction for different temperatures of S. The quantity of SO_3 formed in these experiments only amounted to about 5 per cent. of the

SO_2 . If, however, the SO_2 were only removed after heating the centre tube for a considerable time, as was done in several experiments, about 15 per cent. of the sulphur oxidised over to SO_3 . The addition of nitrogen appeared to have practically no effect on the residual pressure of oxygen, while the rate of the reaction did not vary much as the temperature of the vessel was changed from 100° to 150°C .

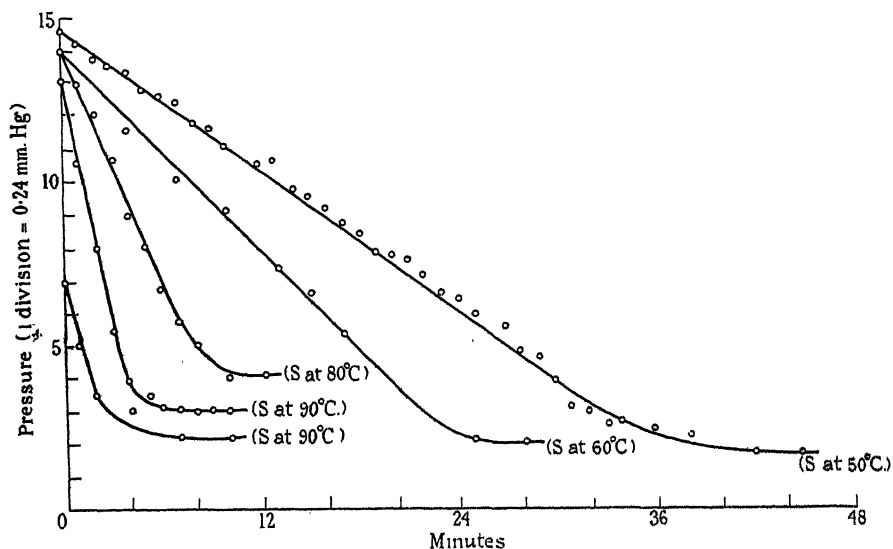


FIG. 5.—Rate of the luminescent reaction and the "residual pressure" found, using the reaction vessel shown in fig. 4.

Discussion.

The investigation shows that sulphur vapour at a pressure of about 0.0025 mm. of mercury, and oxygen at a pressure of about 0.3 mm. of mercury, can react together at a temperature of 100° to 150°C ., provided that the reaction is *started* by a surface at 500°C ., and if the product is removed quickly from the sphere of the reaction. This luminescent reaction appears to be the same as that obtained by Semenoff and Rjabinin when the reaction was initiated by an electric discharge. The fact that there is luminescence, and that the reaction goes spontaneously after the condition which *started* the reaction has been removed, appears to point to a chain reaction. It has been shown by Alyea, Thompson and others* that many "explosions" are initiated on a surface,

* 'Z. phys. Chem.,' B, vol. 10, p. 193 (1930); *ibid.*, vol. 14, p. 359 (1931). cf. Thompson, 'Trans. Faraday Soc.,' vol. 28, p. 412 (1932); and Haber and Oppenheimer, 'Z. phys. Chem.,' B, vol. 16, p. 443 (1932).

and the above experiment may be another example of surface initiation of a chain reaction. A great deal of work still remains to be done on the effect of *wall* temperature on chain initiation, as distinct from the effect of *gas* temperature, before any mechanism of wall initiation can be put forward. In this particular reaction the "wall" requires to be above a temperature of about 450° C. before a chain is started, no appreciable reaction having been noticed when the axial tube was left at 300° C. for half an hour. An interesting point about the reaction is the fact that its rate is independent of the oxygen concentration. This in itself might indicate that it was not a chain reaction, for if it were, the rate of propagation would be proportional to the pressure of the oxygen. If, however, there were reasonable grounds for concluding that in these experiments another factor intervened, causing the rate to be inversely proportional to the oxygen pressure, the two effects would cancel, and the observed result would be independence of oxygen pressure. Such a factor was found: namely, the effect of the oxygen in preventing the removal of SO₂ which the experiments showed acts as an inhibitor of the reaction. When the SO₂ is not removed quickly enough, the reaction stops (unless the "wall" is kept at the high temperature). If this inhibitory effect of SO₂ can be shown to be proportional to the oxygen pressure it will provide the explanation we are seeking. Consideration of the rate of diffusion of SO₂ through the oxygen along the tube leading from the reaction vessel to the liquid air condenser, shows that for an equilibrium condition (*i.e.*, when the amount of SO₂ formed is equal to the amount removed by condensation) the concentration of SO₂ in the reaction vessel must be inversely proportional to the diffusion coefficient of SO₂ into oxygen, and therefore proportional to the pressure of oxygen. The inhibition by SO₂ is therefore proportional to the pressure of oxygen. A further possibility not to be overlooked is that the rate of *initiation* of the chains might also be proportional to the oxygen pressure, but this cannot be so if the above explanation is correct.

Another point is that lowering the temperature of the solid sulphur in the side tube decreases the rate. Semenoff and Rjabinin suggest that the rate of the reaction is proportional to the rate of evaporation of the sulphur. A rough calculation can be made to test whether this assumption is correct. The calculation shows that at 90° C. the change of pressure in 10 minutes should be 9.8 mm. if the rate depended solely on the rate of evaporation, while the rate found experimentally only amounts to 0.3 mm. Although the calculation is admittedly rough, the difference between the values appears to be much greater than should be obtained if Semenoff's assumption were correct. It appears

to us more likely that the rate of the reaction may be affected by the rate of dissociation into S_2 molecules, rather than by rate of evaporation, but we have not been able to develop the idea quantitatively in any useful manner.

The "residual pressure" remains to be discussed. We start the reaction from a surface at 500°C. , and it continues after the hot surface has cooled until the pressure of oxygen has fallen to a certain value, some SO_2 also remaining. A "residual pressure" is usually explained by assuming that the chains are being broken by the wall. This explanation does not appear to be possible here as there is evidence that the chains are not broken in this way. If the chains were broken by the wall we should have to assume that we have present originally active centres which continually initiate chains until the residual pressure is reached. This is apparently not so since as long as the oxygen pressure is kept above the residual pressure by addition of oxygen, the reaction continues, while addition of oxygen after having reached the residual pressure does not cause the reaction to start again. It would, therefore, appear that the active centres which continually initiate the chain must be kept alive by the chains, in other words, the chains are not effectively broken by the walls, a conclusion supported by the observation of Semenov that varying the diameter of the vessel has no effect. One possible explanation is suggested by the fact that the intervention of an open tap, between the reaction vessel and the liquid air trap, stopped the reaction (after the hot surface was allowed to cool). There is evidently a critical diameter and length of this connecting tube, *i.e.*, a critical concentration of SO_2 relative to the oxygen concentration which stops the reaction. As the pressure of oxygen falls, a point may be reached at which the SO_2 is not removed sufficiently rapidly. This would occur when the pressure was such that the mean free path of the molecules was of the same order as the diameter of the tube, and this would be the case at the "residual pressure."

An experiment was also carried out to see whether it was possible to initiate the oxidation photochemically. For this purpose a silica flask of about 250 c.c. capacity was used as the reaction vessel. This flask contained some sulphur and was kept in water at 100°C. , it was connected to the rest of the apparatus and oxygen was introduced. The flask was exposed for 1 hour to a mercury amalgam lamp without any reaction being noticed. The light from an iron spark also failed to initiate the reaction.

We wish to express our thanks to the Carnegie Trustees for the assistance afforded by their scheme for encouraging research, to Professor Kendall for

continued interest in the work, to Dr. H. W. Melville for criticism and advice, and to the Trustees of the Moray Research Fund of Edinburgh University.

Summary.

The low pressure oxidation of sulphur has been studied, particularly the reaction which was found to be initiated on a hot surface and accompanied by a visible glow. The experimental results showed that the observed rate was independent of the observed pressure of oxygen, but this was found to be due to two compensating factors: (1) the rate of oxidation is proportional to the pressure of oxygen, (2) the rate of inhibition by SO_2 is inversely proportional to the pressure of oxygen, hence the net result is independence of oxygen pressure. A residual pressure was observed and a tentative explanation is offered. The effect of foreign gases was not studied in detail, as the one experiment performed indicated that none obtained.

The Internal Conversion Coefficient for Radium C.

By H. R. HULME, Gonville and Caius College, Cambridge.

(Communicated by R. H. Fowler, F.R.S.—Received August 20, 1932.)

§ 1. *Introduction.*—It is well known that the γ -rays emitted from a radioactive nucleus are often partially absorbed by the atomic system, giving rise to secondary β -rays. From observations of the resultant γ -ray intensity, and that of the β -rays, it is possible to infer the proportion of γ -rays reabsorbed in the atomic system. This factor is called the “internal conversion coefficient.” Its theoretical value has been discussed by Miss Swirles† and R. H. Fowler.‡ Miss Swirles treats the nucleus as an oscillating Hertzian doublet, radiating classically, and considers the radiation field as producing photoelectric transitions in the planetary electrons, according to the Schrödinger theory. The rate of emission of γ -rays from the nucleus is taken to be the classical rate of radiation of energy by the dipole, divided by $h\nu$. The values obtained in this way were about 10 times too small, except for the γ -ray of

† ‘Proc. Roy. Soc.,’ A, vol. 116, p. 491 (1927), and vol. 121, p. 447 (1928).

‡ ‘Proc. Roy. Soc.,’ A, vol. 129, p. 1.

energy 14.26×10^5 e.v., which has an internal conversion coefficient several hundred times that given by the theory. This special case has been discussed by Fowler (*loc. cit.*), and we shall not consider it here.

An obvious defect in the theory is the use of Schrödinger's equation, which may not be expected to hold so near the nucleus, or for electrons of such high energy. It therefore seemed possible that the more correct, relativistic equation of Dirac might give results in accordance with experiment in the majority of cases, and the calculation has been carried out by Casimir.† The same model is used, and, for purposes of calculation, the interaction of the other electrons is neglected, so that we have a single electron in the field of a charged nucleus. For the β -rays emitted from the K-shell, we may take the actual nuclear charge in carrying out the calculation. In the case of extremely hard γ -rays, whose energies may be considered large compared with mc^2 , it is legitimate to use the asymptotic expansion for the wave function representing the β -ray. If we apply this theory to the range covered by experiment, we obtain results (Casimir, *loc. cit.*) which are still much too small, so that we were tempted to attribute the bulk of the conversion to some special type of interaction with the nucleus. It seems fairly certain that this must be the case for the γ -ray with $h\nu = 14.26 \times 10^5$ e.v., which has an abnormally high internal conversion coefficient.

At this stage the problem was brought to the notice of the present author by a request from Professor Ehrenfest that Casimir's important conclusions should be checked by an independent calculation. This was done—the formulæ of this paper agree with Casimir's in the limit $h\nu/mc^2 \rightarrow \infty$. But in arranging this check calculation it was found that the calculation could be carried through exactly for all values of $h\nu/mc^2$, subject to a final stage of numerical computation.

In the case of absorption by light elements from a *beam* of γ -rays, calculation shows‡ that we cannot use the asymptotic form of the wave functions for the free electrons in the range covered by experiment, as this gives results which differ from those obtained by using the accurate wave functions. This means that the part of the wave function near the nucleus is important even in the

† 'Phys. Z.,' vol. 32, p. 665 (1931), cf. Gamow, "Atomic Nuclei and Radioactivity," p. 77.

‡ Hulme, 'Proc. Roy. Soc.,' A, vol. 133, p. 381 (1931). Referred to as I. The law connecting the absorption with wave-length varies considerably in the experimental range. Asymptotically we find that the absorption is proportional to the wave-length. See Sauter, 'Ann. Physik,' vol. 11, p. 464 (1931).

case of atoms with small nuclear charge. It therefore seemed worth while to go beyond Casimir's calculation, and to take the actual wave functions instead of the asymptotic expansions, particularly as our present knowledge of the nucleus hardly yet justifies us in attempting any further discussion involving its structure.

The result of the calculation is that the asymptotic formula is indeed insufficient and that *the theoretical internal conversion coefficient is in good agreement with Ellis and Aston's† observed values for certain of the γ -rays of Ra C in the range 500,000 e.v., ($h\nu/mc^2 = 1$), to 1,500,000 e.v., ($h\nu/mc^2 = 3$).* There are other rays in the range for which the calculations do not fit at all. This result has been very carefully checked at all stages as explained in the acknowledgments at the end, and is, we believe, entirely reliable. It seems possible that it may lead to important developments in the theory of the nucleus. A few calculations have also been made of the ratios $K : L_I : L_{II} : L_{III}$ absorptions, which are, for K and L_I , in good agreement with the experimental values.

In this extended calculation we still consider a simplified model obtained by neglecting the effect of the outer shells. We shall neglect the exchange degeneracy and magnetic interaction of the two s -electrons, and calculate the result for one electron in the presence of a nucleus of charge Ze . This must then be multiplied by two to give the internal conversion coefficient for the two electrons of the K-shell. Following Miss Swirles and Casimir, we shall place an oscillating dipole at the nucleus, to represent the mechanism emitting γ -rays, and the first part of the calculation is very similar to that of these two authors.

§2. *Perturbation Theory.*—We shall use the notation of I, writing the wave equation of the electron as

$$\left[\frac{E}{c} + \frac{e}{c} A_0 + \rho_1 \left(\boldsymbol{\sigma}, \mathbf{p} + \frac{e}{c} \mathbf{A} \right) + \rho_3 mc \right] \Psi = 0, \quad (1)$$

where E is the energy of the system, A_0 and \mathbf{A} the scalar and vector potentials, \mathbf{p} the momentum vector (p_x, p_y, p_z) and $\boldsymbol{\sigma}$ the vector ($\sigma_x, \sigma_y, \sigma_z$), and $e < 0$. The quantities $\sigma_x, \sigma_y, \sigma_z, \rho_1, (\rho_2)$ and ρ_3 do not commute with each other and may be conveniently represented by certain matrices of four rows and columns which obey the same non-commutability relations. In this representation the wave function Ψ has four components.

† 'Proc. Roy. Soc.,' A, vol. 129, p. 180 (1930).

Let us take as the unperturbed system an electron under the influence of a central charge Ze . We have then $A = 0$ and $A_0 = Ze/r$. The perturbing potentials may be found from the Hertzian vector of the doublet which we shall take to be $\pi_z + \pi_z^*$, where

$$\pi_z = B_0 e \cdot \exp(-2\pi i \nu t + iqr)/iqr,$$

$q = 2\pi\nu/c$ and the asterisk indicates that the conjugate complex value be taken. This yields for the perturbing potentials the following values:—

$$\left. \begin{aligned} A_0 &= -B_0 e \cdot \exp(-2\pi i \nu t + iqr) \cos \theta \left\{ \frac{1}{r} - \frac{1}{iqr^2} \right\} + \text{conjugate complex} \\ A_z &= -B_0 e \cdot \exp(-2\pi i \nu t + iqr) \frac{1}{r} + \text{conjugate complex} \\ A_x &= A_y = 0, \end{aligned} \right\} \quad (2)$$

where θ is measured from the z -axis.

We may omit the conjugate complex part in the calculation of the transition probabilities, since it is only important for transitions where a γ -ray is emitted. (If $h\nu'$ denote the increase of energy of the system, this part of the perturbation occurs with $\nu + \nu'$ in the denominator, which is therefore always very large in the case of absorption.) Treating the Hertzian oscillator as a classical system we find that the amount of energy radiated per unit time is $4B_0^2 e^2 c q^2 / 3$ ergs.[†] or $8\pi B_0^2 e^2 q / 3h$ quanta.

Suppose we have solved equation (1) for the undisturbed model atom, and let ψ_0 represent the normalised wave function for the ground state, and ψ_k that of a possible final state, where the electron is “free,” the time factor being omitted in both cases. If ψ_k be normalised so that its components represent one electron entering or leaving a large sphere, with centre at the origin, per unit time, then we find in the usual way a value for the transition probability, per quantum of γ -radiation, given by

$$\Sigma \frac{3h}{8\pi q B_0^2 e^2} \left(\frac{2\pi}{h} \right)^2 |(\psi_k | -eA_0 - eA_x \sigma_1 \sigma_z | \psi_0)|^2, \quad (3)$$

where we must sum over all the possible final states, and the time factor has now disappeared from A_0 and A_z .

§3. *Calculation of the Matrix Elements.*—In this calculation we shall take the axis of the dipole along the z -axis, and we shall assume that the atomic system

[†] See, for example, Abraham, “Theorie der Electrizität,” vol. 2, § 8.

is quantised about this axis. In the calculations made by Miss Swirles, the axis of quantisation was fixed and then the results averaged over all directions of the dipole. For an s -state we may easily verify that these two procedures give the same result, as they must do for any spherically symmetrical state. (The averaging introduces a factor of one-third, which is compensated for by the fact that there are more possible final states for an arbitrary direction of the dipole.)

The explicit form of the wave functions depends upon the matrices chosen to represent the σ and ρ 's in (1), and in the following we shall use those given by Dirac. The solutions are then found to be of two types†

$$\left. \begin{aligned} \psi_1 &= -iF_k P_{k+1}^u & \psi_2 &= -iF_k P_{k+1}^{u+1} \\ \psi_3 &= (k+u+1) G_k P_k^u & \psi_4 &= (-k+u) G_k P_k^{u+1} \end{aligned} \right\} \quad (4A)$$

and

$$\left. \begin{aligned} \psi_1 &= -i(k+u) F_{-k-1} P_{k-1}^u & \psi_2 &= -i(-k+u+1) F_{-k-1} P_{k-1}^{u+1} \\ \psi_3 &= G_{-k-1} P_k^u & \psi_4 &= G_{-k-1} P_k^{u+1} \end{aligned} \right\}, \quad (4B)$$

where F_k and G_k satisfy

$$\left. \begin{aligned} \left(A^2 + \frac{\gamma}{r} \right) F_k + \frac{dG_k}{dr} - \frac{k}{r} G_k &= 0 \\ \left(B^2 - \frac{\gamma}{r} \right) G_k + \frac{dF_k}{dr} + \frac{k+2}{r} F_k &= 0 \end{aligned} \right\}, \quad (5)$$

with

$$\begin{aligned} A^2 &= \frac{2\pi}{h} \left(mc + \frac{E}{c} \right) \\ B^2 &= \frac{2\pi}{h} \left(mc - \frac{E}{c} \right) \end{aligned} \quad \gamma = \frac{2\pi e^2 Z}{ch} = \frac{Z}{137} \text{ approximately,}$$

and P_k^u is the associated Legendre function given by

$$P_k^u = (k-u)! \sin^u \theta \left(\frac{d}{d \cos \theta} \right)^{k+u} \frac{(\cos^2 \theta - 1)^k}{2^k k!} e^{iu\phi},$$

u and k being any numbers such that the Legendre functions involved have a meaning. In this representation the axis of quantisation is the z -axis. When $E < mc^2$, suitable solutions of (5) can be obtained in the form of polynomials

† Darwin, 'Proc. Roy. Soc.,' A, vol. 118, p. 654 (1928).

for a set of discrete values of the energy. The two states of lowest energy, corresponding to the two s -orbits, are both of type (4A) and are given by :

$$\left. \begin{aligned} \psi_1 &= -\frac{i\gamma}{1 + \sqrt{1 - \gamma^2}} r^\beta e^{-r/a_0} \cos \theta \\ \psi_2 &= -\frac{i\gamma}{1 + \sqrt{1 - \gamma^2}} r^\beta e^{-r/a_0} \sin \theta e^{i\phi} \\ \psi_3 &= r^\beta e^{-r/a_0} \quad \psi_4 = 0 \end{aligned} \right\}, \quad (6)$$

for which $k = 0$ and $u = 0$ and

$$\left. \begin{aligned} \psi_1 &= \frac{i\gamma}{1 + \sqrt{1 - \gamma^2}} r^\beta e^{-r/a_0} \sin \theta e^{-i\phi} \\ \psi_2 &= -\frac{i\gamma}{1 + \sqrt{1 - \gamma^2}} r^\beta e^{-r/a_0} \cos \theta \\ \psi_3 &= 0 \quad \psi_4 = -r^\beta e^{-r/a_0} \end{aligned} \right\}, \quad (7)$$

for which $k = 0$ and $u = -1$ and we have put

$$a_0 = \frac{\hbar^2}{4\pi^2 m Z e^2}, \quad \beta = \sqrt{1 - \gamma^2} - 1.$$

The energies of both these states are equal to $mc^2 \sqrt{1 - \gamma^2}$ and the normalisation factor is given by

$$[\xi(E_0)]^2 4\pi \left(\frac{a_0}{2}\right)^{1+2\sqrt{1-\gamma^2}} \Gamma\{1 + 2\sqrt{1 - \gamma^2}\} \frac{2}{1 + \sqrt{1 - \gamma^2}} = 1. \quad (8)$$

Any positive value of $(E - mc^2)$ will yield a permissible solution of (5). We require a normalised solution valid for all r , and this may be obtained in the form of a contour integral. In this case, B is a pure imaginary, and if we put

$$\mathfrak{G}_k = A F_k + i |B| G_k \quad (9)$$

we find (Hulme, *loc. cit.*),

$$\mathfrak{G}_k = [(k - s) - i(b + c)] r^s a^{2s+2} e^{-\pi b} \int_{-1}^{+1} (1 - u)^{s-ib+1} (1 + u)^{s+ib} e^{i a r u} du \quad (10A)$$

and

$$\mathfrak{G}_{-k-1} = [(b - c) + i(k - 1 - s')] r^{s'} a^{2s'+2} e^{-\pi b} \int_{-1}^{+1} (1 - u)^{s'-ib+1} (1 + u)^{s'+ib} e^{i a r u} du, \quad (10B)$$

where

$$\left. \begin{aligned} AB &= ia, \\ \frac{\gamma}{2} \left(\frac{A}{B} - \frac{B}{A} \right) &= -ib, \\ \frac{\gamma}{2} \left(\frac{A}{B} + \frac{B}{A} \right) &= -ic, \end{aligned} \right\} \begin{aligned} s &= \sqrt{\{(k+1)^2 - \gamma^2\} - 1}, \\ s' &= \sqrt{\{k^2 - \gamma^2\} - 1} \end{aligned} \quad (11)$$

We shall express the normalising factor as a product of $\xi(E, k)$, the normalising factor for the radial part of the wave function, and $\xi(k, u)$, that for the angular part. We obtain for the solutions (4A)

$$\left. \begin{aligned} \xi(E, k) &= \left(\frac{2\pi E}{\hbar c^2 a} \right)^{\frac{1}{2}} \frac{A |B|}{(A^2 + |B|^2)^{\frac{1}{2}}} K^{-1} \\ \text{with} \quad K &= \frac{1}{2} \sqrt{\{(k-s)^2 + (b+c)^2\}} |\Gamma(s+ib+1)| e^{-3\pi b/2} (2a)^{s+1} \\ \text{and} \quad \xi(k, u) &= \{4\pi (k+u+1)! (k-u)!\}^{-\frac{1}{2}} \end{aligned} \right\} \quad (12)$$

To obtain the corresponding normalising factors for the solutions (4B) we replace k by $k-1$, s by s' and c by $-c$ in the above.

Let us now consider the values of the matrix elements in (3). For the initial state we should really take a combination of the states given by (6) and (7) with arbitrary phase factors, and then average the results over all possible values of the phase factors. The two states represent the spin pointing in opposite directions along the z -axis, and it is easily seen from the symmetry of the perturbation that they are entirely equivalent physically, and will yield the same result. We shall therefore only consider the first state.

With the representation used we find

$$\rho_{1\sigma_z} = \begin{Bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{Bmatrix}.$$

Since we are dealing with radiation from a dipole, we should expect some accurate selection rule, valid for all values of ν . We find in fact that only two final states are possible, namely those given by

$$\Delta k = 1, \quad \Delta u = 0.$$

They are therefore $k=1$, $u=0$, of types (4A) and (4B). Consider first the

transition to the state of type (4A). Neglecting, for the moment, the normalisation factors, we have

$$\sum_{s=1}^3 \sum_{k=1}^4 \psi_s^* \psi_s = 2 \cos \theta [G_k^* + \gamma \{1 + \sqrt{1 - \gamma^2}\}^{-1} F_k^*] r^\beta e^{-r/a_0},$$

$$\psi_k^* \psi_1 - \psi_k^* \psi_2 + \psi_k^* \psi_3 - \psi_k^* \psi_4$$

$$= -i [\gamma \{1 + \sqrt{1 - \gamma^2}\}^{-1} (1 + \cos^2 \theta) G_k^* + (1 - 3 \cos^2 \theta) F_k^*] r^\beta e^{-r/a_0}.$$

For the corresponding matrix element in (3), after dividing by $B_0^2 e^2$, we find a value given by

$$\int_0^\pi \int_0^{2\pi} \int_0^\infty \left[G_k \{2(\beta + 2) \cos^2 \theta - i\gamma (1 + \cos^2 \theta)\} \right. \\ \left. + F_k \{2\gamma \cos^2 \theta - i(\beta + 2)(1 - 3 \cos^2 \theta)\} \right. \\ \left. + \frac{i}{qr} 2 \cos^2 \theta \{(\beta + 2) G_k + \gamma F_k\} \right] \\ \times (\beta + 2)^{-1} e^{iqr} r^{\beta+1} e^{-r/a_0} \sin \theta d\theta d\phi dr,$$

where we have put $G_k^* = G_k$, since it is real.

On integrating with respect to θ and ϕ we obtain the expression

$$\frac{8\pi}{3} \int_0^\infty \left[G_k \{(\beta + 2) - 2i\gamma\} + \gamma F_k + \frac{i}{qr} \{(\beta + 2) G_k + \gamma F_k\} \right] \\ \times (\beta + 2)^{-1} r^{\beta+1} e^{iqr - r/a_0} dr. \quad (13A)$$

Similarly, for the transition to the state of type (4B) we obtain for the corresponding expression the value

$$\frac{4\pi}{3} \int_0^\infty \left[G_{-k-1} \{(\beta + 2) + i\gamma\} + F_{-k-1} \{3i(\beta + 2) + \gamma\} \right. \\ \left. + \frac{i}{qr} \{(\beta + 2) G_{-k-1} + \gamma F_{-k-1}\} \right] (\beta + 2)^{-1} r^{\beta+1} e^{iqr - r/a_0} dr. \quad (13B)$$

We now express F_k and G_k in terms of \mathcal{G}_k and \mathcal{G}_k^* by means of equation (9), remembering that B is a pure imaginary. We find for the expressions (13A) and (13B) the following values,

$$\frac{4\pi}{3} \int_0^\infty \left[\mathcal{G}_k \left\{ \gamma \left(\frac{1}{A} - \frac{2}{|B|} \right) - i \frac{\beta + 2}{|B|} \right\} + \mathcal{G}_k^* \left\{ \gamma \left(\frac{1}{A} + \frac{2}{|B|} \right) + i \frac{\beta + 2}{|B|} \right\} \right. \\ \left. + \frac{i}{qr} \left\{ \mathcal{G}_k \left(\frac{\gamma}{A} - i \frac{\beta + 2}{|B|} \right) + \mathcal{G}_k^* \left(\frac{\gamma}{A} + i \frac{\beta + 2}{|B|} \right) \right\} \right] (\beta + 2)^{-1} e^{iqr - r/a_0} r^{\beta+1} dr, \quad (14A)$$

and

$$\begin{aligned} & \frac{2\pi}{3} \int_0^\infty \left[\mathfrak{G}_{-k-1} \left\{ \frac{\gamma + 3i(\beta + 2)}{A} + \frac{\gamma - i(\beta + 2)}{|B|} \right\} \right. \\ & + \mathfrak{G}_{-k-1}^* \left\{ \frac{\gamma + 3i(\beta + 2)}{A} - \frac{\gamma - i(\beta + 2)}{|B|} \right\} + \frac{i}{qr} \left\{ \mathfrak{G}_{-k-1} \left(\frac{\gamma}{A} - i \frac{\beta + 2}{|B|} \right) \right. \\ & \left. \left. + \mathfrak{G}_{-k-1}^* \left(\frac{\gamma}{A} + i \frac{\beta + 2}{|B|} \right) \right\} \right] (\beta + 2)^{-1} e^{iar-r/a_0} r^{\beta+1} dr. \end{aligned} \quad (14B)$$

§ 4. *Evaluation of the Integrals.*—The integrals occurring above cannot be evaluated in finite terms, but we can express them in terms of hypergeometric series, which can then be evaluated numerically. Consider first the integral

$$\int_0^\infty \mathfrak{G}_k e^{iar-r/a_0} r^{\beta+1} dr, \quad (15)$$

where \mathfrak{G}_k is given by (10A). Substituting for \mathfrak{G}_k we obtain

$$\begin{aligned} & \{(k-s) - i(b+c)\} a^{2s+2} e^{-\pi b} \int_0^\infty e^{iar-r/a_0} r^{s+\beta+1} \\ & \times \int_{-1}^{+1} (1-u)^{s-i b+1} (1+u)^{s+i b} e^{iaru} du dr. \end{aligned} \quad (16)$$

Now if E_0 and E be the energies of the initial and final states of the system we have

$$\begin{aligned} E &= E_0 + h\nu \\ &= mc^2 \sqrt{1-\gamma^2} + hqc/2\pi. \end{aligned}$$

Also from (11) we have

$$a^2 = \frac{4\pi^2}{c^2 h^2} \{ (E + mc^2) (E - mc^2) \},$$

which gives, putting $\theta = mc^2/h\nu$ and $\beta = \sqrt{1-\gamma^2} - 1$,

$$\begin{aligned} a^2 &= q^2 \{ 1 + (2\beta + 2) \theta + \beta(\beta + 2) \theta^2 \} \\ &= q^2 / \tau^2, \text{ say.} \end{aligned} \quad (17)$$

Further, if we put $1/a_0 = \delta a$ we may change the variable of integration in (16) by writing r instead of ar , obtaining

$$\begin{aligned} & \{(k-s) - i(b+c)\} e^{-\pi b} a^{s-\beta} \\ & \times \int_0^\infty \int_{-1}^{+1} e^{-(\delta-i\tau-iu)r} r^{\beta+s+1} (1-u)^{s-i b+1} (1+u)^{s+i b} dr du. \end{aligned}$$

Integrating this with respect to r we find

$$\{(k-s) - i(b+c)\} e^{-\pi b} a^{s-\beta} \Gamma(s+\beta+2) \int_{-1}^{+1} \frac{(1-u)^{s-ib+1} (1+u)^{s+ib}}{(\delta - i\tau - iu)^{s+\beta+2}} du.$$

Writing M for the expression outside the integral, and putting $1+u=2u'$, the expression becomes

$$\frac{M2^{2s+2}}{[\delta - i(\tau-1)]^{s+\beta+2}} \int_0^1 (1-u')^{s-ib+1} u'^{s+ib} (1-zu')^{-(s+\beta+2)} du',$$

where $z=2/(1-\tau-ib)$. The integral is now expressible as a hypergeometric function,† and we obtain for (15) the value

$$\frac{M2^{2s+2}}{[\delta - i(\tau-1)]^{s+\beta+2}} \frac{\Gamma(s+ib+1) \Gamma(s-ib+2)}{\Gamma(2s+3)} \times F(s+\beta+2, s+ib+1; 2s+3; z). \quad (18)$$

For γ -rays, z is large, and it is convenient to evaluate the hypergeometric function by transforming it into a function of z^{-1} .

We use for formula‡

$$\begin{aligned} & \frac{\Gamma(a) \Gamma(b)}{\Gamma(c)} F(a, b; c; z) \\ &= \frac{\Gamma(a) \Gamma(b-a)}{\Gamma(c-a)} (-z)^{-a} F(a, 1-c+a; 1-b+a; z^{-1}) \\ &+ \frac{\Gamma(b) \Gamma(a-b)}{\Gamma(c-b)} (-z)^{-b} F(b, 1-c+b; 1-a+b; z^{-1}), \quad (19) \end{aligned}$$

where $\arg(-z)$ is so chosen that $|\arg(-z)| < \pi$. In this way we obtain a value for the integral (15) given by

$$\begin{aligned} & [(k-s) - i(b+c)] \cdot L \cdot \left[F(s+ib+1, ib-s-1; ib-\beta; z^{-1}) \right. \\ &+ \frac{\Gamma(s+\beta+2) \Gamma(ib-\beta-1) \Gamma(s+2-ib)}{\Gamma(s+1-\beta) \Gamma(s+ib+1) \Gamma(\beta+1-ib)} (-z)^{-(s+1-ib)} \\ &\left. \times F(s+\beta+2, \beta-s; \beta+2-ib; z^{-1}) \right] \quad (20) \end{aligned}$$

where

$$L = \frac{e^{-3\pi b/2} a^{s-\beta} 2^{s+1-ib} \Gamma(s+ib+1) \Gamma(\beta+1-ib) i^{s+1}}{[\delta - i(\tau-1)]^{s+1-ib}}.$$

† Whittaker and Watson, "Modern Analysis," ch. 14.

‡ Barnes, 'Proc. Lond. Math. Soc.' vol. 6, p. 141 (1908). The formula is given wrongly in Whittaker and Watson.

Consider now the integral involving \mathfrak{G}_k^* —

$$\int_0^\infty \mathfrak{G}_k^* e^{iqr-\tau/a_0} r^{\beta+1} dr. \quad (21)$$

The *conjugate* of this expression is obtained by replacing q by $-q$ in (15). This leads to an expression equal to (18) with τ replaced by $-\tau$ and z by $z' = 2/(1 + \tau - i\delta)$. We transform the hypergeometric function by the formula

$$\mathbf{F}(a, b; c; z') = (1 - z')^{-a} \mathbf{F}(a, c - b; c; z'/(z' - 1)).$$

We find $z'/(z' - 1) = z^*$, so that on taking the conjugate complex again, we obtain for the integral involving \mathfrak{G}_k^* , an expression in z , similar to (16). On transforming as before, this gives for the integral (21) the value

$$\frac{M^* 2^{2s+2}}{[\delta - i(\tau - 1)]^{s+\beta+2}} \frac{\Gamma(s + ib + 2) \Gamma(s - ib + 1)}{\Gamma(2s + 3)} \times \mathbf{F}(s + \beta + 2, s + ib + 2; 2s + 3; z), \quad (22)$$

and finally,

$$\begin{aligned} & [(k - s) + i(b + c)] \cdot \mathbf{L} \cdot \left[\frac{(s + ib + 1)}{(\beta - ib)} (-z)^{-1} \right. \\ & \quad \times \mathbf{F}(s + ib + 2, ib - s, ib + 1 - \beta; z^{-1}) \\ & \quad + \frac{\Gamma(s + \beta + 2) \Gamma(ib - \beta) \Gamma(s - ib + 1)}{\Gamma(s + 1 - \beta) \Gamma(s + ib + 1) \Gamma(\beta + 1 - ib)} (-z)^{-(\beta+1-ib)} \\ & \quad \left. \times \mathbf{F}(s + \beta + 2, \beta - s; \beta + 1 - ib; z^{-1}) \right]. \quad (23) \end{aligned}$$

In the expression (14A) there occur also the integrals

$$\frac{1}{q} \int_0^\infty \mathfrak{G}_k e^{iqr-\tau/a_0} r^\beta dr \quad (24)$$

and

$$\frac{1}{q} \int_0^\infty \mathfrak{G}_k^* e^{iqr-\tau/a_0} r^\beta dr. \quad (25)$$

The values of these may be obtained from the expressions (20) and (23) by replacing β by $\beta - 1$. For the purposes of numerical calculation, however, it is preferable to express them directly in terms of the two integrals already calculated. This may be done by means of the relations existing between contiguous hypergeometric functions as follows. Let us call the integrals

(15), (21), (24) and (25), P, Q, R and S respectively. Replacing β by $\beta - 1$ in the expressions (18) and (22) for P and Q, we obtain values for R and S.

$$R = \frac{a}{q} \frac{M 2^{2s+2}}{(s + \beta + 1) [\delta - i(\tau - 1)]^{\beta+s+1}} \frac{\Gamma(s + i\bar{b} + 1) \Gamma(s - i\bar{b} + 2)}{\Gamma(2s + 3)} \\ \times F(s + \beta + 1, s + i\bar{b} + 1; 2s + 3; z)$$

$$S = \frac{a}{q} \frac{M^* 2^{2s+2}}{(s + \beta + 1) [\delta - i(\tau - 1)]^{\beta+s+1}} \frac{\Gamma(s + i\bar{b} + 2) \Gamma(s - i\bar{b} + 1)}{\Gamma(2s + 3)} \\ \times F(s + \beta + 1, s + i\bar{b} + 2; 2s + 3; z).$$

We now use the two recurrence formulæ

$$\left. \begin{aligned} (c - a) F(a - 1, b; c; z) &= b(1 - z) F(a, b + 1; c; z) \\ &\quad + (c - b - a) F(a, b; c; z) \\ \text{and} \\ (c - a) F(a - 1, b + 1; c; z) &= (b - a + 1)(1 - z) F(a, b + 1; c; z) \\ &\quad + (c - b - 1) F(a, b; c; z) \end{aligned} \right\} \quad (26)$$

which enable us to express the hypergeometric functions in R and S in terms of those in P and Q. Substituting in these relations we obtain

$$\left. \begin{aligned} \frac{q}{a} [\delta - i(\tau - 1)]^{-1} (s + \beta + 1) (s - \beta + 1) M^* \cdot R &= (-i\bar{b} - \beta) M^* \cdot P \\ &\quad + (s - i\bar{b} + 1)(1 - z) M \cdot Q \\ \text{and} \\ \frac{q}{a} [\delta - i(\tau - 1)]^{-1} (s + \beta + 1) (s - \beta + 1) M \cdot S &= (s + i\bar{b} + 1) M^* \cdot P \\ &\quad + (i\bar{b} - \beta)(1 - z) M \cdot Q. \end{aligned} \right\} \quad (27)$$

These two equations enable us to calculate the numerical values of R and S when we know those of P and Q.

When the final state is of type (4B) all the formulæ of this section hold, provided we replace s by s' and $[(k - s) - i(b + c)]$ by $[(b - c) + i(k - 1 - s')]$. We shall call the new values of L and M, L' and M'. In this case a further simplification is possible if we observe that $s' = \beta$ when $k = 1$. We can, in fact, find a relation between the integrals P and Q, which reduces the numerical work. The relation

$$(c - b) F(a, b; c + 1; z) + b F(a, b + 1; c + 1; z) = c F(a, b; c; z)$$

gives

$$\begin{aligned} (s' - i\bar{b} + 1) F(\beta + s' + 2, s' + i\bar{b} + 1; 2s' + 3; z) \\ + (s' + i\bar{b} + 1) F(\beta + s' + 2, s' + i\bar{b} + 2; 2s' + 3; z) \\ = (2s' + 2) F(\beta + s' + 2, s' + i\bar{b} + 1; 2s' + 2; z). \end{aligned}$$

If $s' = \beta$, the last hypergeometric function reduces to

$$(1 - z)^{-(\beta + ib + 1)},$$

and from (18) and (22) we have

$$M^* P' + M' Q' = \frac{M' \cdot M'^* \cdot 2^{2\beta+2}}{[\delta - i(\tau - 1)]^{2\beta+2}} \frac{\Gamma(\beta + ib + 1) \Gamma(\beta - ib + 1)}{\Gamma(2\beta + 2)} (1 - z)^{-(\beta + ib + 1)}, \quad (28)$$

which is the required equation.

The expressions obtained for the integrals P, Q, etc., have still to be multiplied by the normalising factors for the initial and final states. These are given by (8) and (12). The internal conversion coefficient is then found from the formula (2) and the expressions (14A) and (14B).

The final result is rather cumbersome and we shall express it as follows. The integral (15) is given by the expression (20) which we shall write as

$$|(k - s) - i(b + c)| \cdot L \cdot I_1.$$

Similarly for the expression (23) we write

$$|(k - s) - i(b + c)| \cdot L \cdot I_2.$$

The integrals (24) and (25) are found in terms of these two, and we shall put them equal to

$$|(k - s) - i(b + c)| \cdot L \cdot I_3$$

and

$$|(k - s) - i(b + c)| \cdot L \cdot I_4.$$

When the final state is of type (4A) the integrals are

$$|(b - c) + i(k - 1 - s')| \cdot L' \cdot I'_1, \text{ etc.}$$

These are now substituted in (14A) and (14B) which are to be multiplied by the normalisation factors for the initial and final states, and substituted into the expression (3) for the internal conversion coefficient. After some algebra we obtain the final result

$$I_K = \frac{1}{24} \frac{\gamma b}{Z} \theta \frac{(2\gamma)^{2\beta+2}}{\Gamma(2\beta + 3) \cdot (\beta + 2)} \times \frac{|\Gamma(\beta + 1 - ib)|^2}{\left| \left\{ \gamma + \frac{i}{\theta} \left(\frac{a}{q} - 1 \right) \right\}^{2(\beta+1-ib)} \right|} \frac{2A^2 |B|^2}{A^2 + |B|^2} (2|\mathfrak{U}|^2 + |\mathfrak{B}|^2), \quad (29)$$

where

$$\begin{aligned} \mathfrak{U} = & I_1 \left\{ \gamma \left(\frac{1}{A} - \frac{2}{|B|} \right) - i \frac{\beta + 2}{|B|} \right\} + I_2 \left\{ \gamma \left(\frac{1}{A} + \frac{2}{|B|} \right) + i \frac{\beta + 2}{|B|} \right\} \\ & + I_3 \left\{ \frac{\beta + 2}{|B|} + \frac{i\gamma}{A} \right\} + I_4 \left\{ - \frac{\beta + 2}{|B|} + \frac{i\gamma}{A} \right\} \end{aligned}$$

and

$$\begin{aligned}\mathfrak{B} = & I'_1 \left\{ \gamma \left(\frac{1}{A} + \frac{1}{|B|} \right) + i(\beta + 2) \left(\frac{3}{A} - \frac{1}{|B|} \right) \right\} \\ & + I'_2 \left\{ \gamma \left(\frac{1}{A} - \frac{1}{|B|} \right) + i(\beta + 2) \left(\frac{3}{A} + \frac{1}{|B|} \right) \right\} \\ & + I'_3 \left\{ \frac{\beta + 2}{|B|} + \frac{i\gamma}{A} \right\} + I'_4 \left\{ -\frac{\beta + 2}{|B|} + \frac{i\gamma}{A} \right\}.\end{aligned}$$

This result is obtained for one electron, and, before comparing it with experiment, we must multiply it by two to allow for the two electrons of the K-shell.

From the above expression we can easily obtain a formula giving the value of I_k as $\lambda \rightarrow 0$. This limit implies the following ones:—

$$\begin{aligned}\theta &\rightarrow 0, & z &\rightarrow \infty \\ \tau &\rightarrow 1, & q &\rightarrow a \rightarrow \infty\end{aligned}$$

From expressions (20) and (23) we have

$$|I_1| \rightarrow 1, \quad I_2 \rightarrow 0.$$

We then see from (27) that R and $S \rightarrow 0$, so that I_3 and $I_4 \rightarrow 0$.

Similar results hold for I'_1 , etc. Also as $\theta \rightarrow 0$, $A/|B| \rightarrow 1$, and $b \rightarrow \gamma$, so that

$$\left\{ \gamma + \frac{i}{\theta} \left(\frac{a}{q} - 1 \right) \right\} \rightarrow \gamma + i(\beta + 1)$$

and

$$\begin{aligned}|\mathfrak{A}|^2 &\rightarrow 2(\beta + 2)/A^2 \\ |\mathfrak{B}|^2 &\rightarrow 8(\beta + 2)/A^2.\end{aligned}$$

Putting these values in we obtain a value for the internal conversion coefficient per electron given by

$$I_{K, \text{Asymptotic}} = \frac{1}{2} \frac{\gamma^2 \theta}{Z} (2\gamma)^{2\sqrt{(1-\gamma^2)}} \frac{|\Gamma\{\sqrt{(1-\gamma^2)} - i\gamma\}|^2}{\Gamma\{1 + 2\sqrt{(1-\gamma^2)}\}} e^{-2\gamma \arccos \gamma}, \quad (30)$$

which is the formula given by Casimir.†

§ 5. *Internal Conversion by the Two L_I Electrons.*—Experiment shows that the two K-electrons produce the strongest lines in the β -ray spectrum, the next strongest being due to absorption of γ -rays by the two L_I electrons. The

† Casimir, *loc. cit.* Dr. Casimir has pointed out that by using the property

$$\psi = \psi_{\text{Asymptotic}} \{1 + O(1/ar)\},$$

we may easily prove that the use of the asymptotic wave functions will give a correct result in this problem if $h\nu \gg mc^2$.

amount of this absorption may be found exactly as for the K-electrons, the algebra, however, is a little more complicated.

For discrete states the values of F_k and G_k have been given by Darwin (*loc. cit.*)

$$\left. \begin{aligned} F_k &= \frac{\gamma}{N+k'+n'} e^{-r/a_0 N} \\ &\times \left\{ r^{k'+n'-1} (N+k+1) - r^{k'+n'-2} a_0 N (N+k+2) \frac{n'(n'+2k')}{2} \right. \\ &\quad \left. + r^{k'+n'-3} (a_0 N)^2 (N+k+3) \frac{n'(n'-1)(n'+2k')(n'+2k'-1)}{2 \cdot 4} - \dots \right\} \\ G_k &= e^{-r/a_0 N} \left\{ r^{k'+n'-1} (N+k+1) - r^{k'+n'-2} a_0 N (N+k) \frac{n'(n'+2k')}{2} \right. \\ &\quad \left. + r^{k'+n'-3} (a_0 N)^2 (N+k-1) \frac{n'(n'-1)(n'+2k')(n'+2k'-1)}{2 \cdot 4} - \dots \right\} \end{aligned} \right\} \quad (31)$$

with

$$\left. \begin{aligned} k' &= \sqrt{\{(k+1)^2 - \gamma^2\}} \\ \text{and} \quad N &= \sqrt{\{(k' + n')^2 + \gamma^2\}}, \end{aligned} \right\} \quad (32)$$

where n' is an integer. The series terminate and the corresponding energy is given by

$$E = mc^2 (k' + n')/N. \quad (33)$$

For F_{-k-1} and G_{-k-1} we must replace k by $-k-1$ in the above. These wave functions are only valid for one electron in the field of a fixed charge Ze . We shall therefore neglect the effect of all the other electrons and assume these wave functions represent the L_I electrons when $n' = 1$, $k = 0$ and Z is equal to the nuclear charge. Since $k = 0$ the wave functions for the two L_I -electrons differ from those of the K-electrons only in the radial part. The possible initial states are therefore given by $u = 0$, $u = -1$, and are both of type (4A). As before they are both spherically symmetrical and yield the same result. We shall consider the first state only, keeping the direction of the nuclear dipole fixed along the z -axis. Putting in $n' = 1$, $k = 0$ and $N^2 = 2(k' + 1)$, the equations (31) become

$$\left. \begin{aligned} F_k &= \gamma (N+1) a_0 e^{-r/N a_0} \{N \delta_1 a r - N^2 (N-1)/2\} = \gamma (N+1) a_0 e^{-r/N a_0} f, \\ G_k &= (N+1) a_0 e^{-r/N a_0} \{2 \delta_1 a r / (N+2) - (N-1)\} = \gamma (N+1) a_0 e^{-r/N a_0} g, \end{aligned} \right\} \quad (34)$$

where we have put $N a a_0 = 1/\delta_1$.

The wave functions for the initial state are normalised in the usual way and we find

$$[\xi(E_0)]^2 4\pi \left(\frac{a_0 N}{2}\right)^{2\beta+5} \Gamma(N^2) \frac{8N(N+1)}{N+2} = 1.$$

The calculations follow those for the K-shell very closely, and, corresponding to (14A) and (14B), we obtain the following expressions

$$\begin{aligned} \frac{4\pi}{3} \int_0^\infty \left[\mathfrak{G}_k \left\{ \gamma \cdot f \cdot \left(\frac{1}{A} - \frac{2}{|B|} \right) - \frac{ig}{|B|} \right\} + \mathfrak{G}_k^* \left\{ \gamma \cdot f \cdot \left(\frac{1}{A} + \frac{2}{|B|} \right) + \frac{ig}{|B|} \right\} \right. \\ \left. + \frac{i}{qr} \left\{ \mathfrak{G}_k \left(\frac{\gamma \cdot f}{A} - \frac{ig}{|B|} \right) + \mathfrak{G}_k^* \left(\frac{\gamma \cdot f}{A} + \frac{ig}{|B|} \right) \right\} \right] (N+1) a_0 r^{\beta+1} e^{iqr-r/N a_0} dr, \end{aligned} \quad (35A)$$

and

$$\begin{aligned} \frac{2\pi}{3} \int_0^\infty \left[\mathfrak{G}_{-k-1} \left\{ \gamma \cdot f \cdot \left(\frac{1}{A} + \frac{1}{|B|} \right) + ig \left(\frac{3}{A} - \frac{1}{|B|} \right) \right\} \right. \\ \left. + \mathfrak{G}_{-k-1}^* \left\{ \gamma \cdot f \cdot \left(\frac{1}{A} - \frac{1}{|B|} \right) + ig \left(\frac{3}{A} + \frac{1}{|B|} \right) \right\} + \frac{i}{qr} \left\{ \mathfrak{G}_{-k-1} \left(\frac{\gamma \cdot f}{A} - \frac{ig}{|B|} \right) \right. \right. \\ \left. \left. + \mathfrak{G}_{-k-1}^* \left(\frac{\gamma \cdot f}{A} + \frac{ig}{|B|} \right) \right\} \right] (N+1) a_0 r^{\beta+1} e^{iqr-r/N a_0} dr. \end{aligned} \quad (35B)$$

On putting in the values of f and g these become

$$\begin{aligned} \frac{4\pi}{3} \int_0^\infty \left[\mathfrak{G}_k \left\{ \gamma(N-1) \left(\frac{1}{A} - \frac{2}{|B|} \right) - \frac{a\delta_1}{q} \frac{N}{|B|} - \frac{iN^2(N-1)}{2|B|} - \frac{ia\delta_1}{q} \frac{2}{N+2} \frac{\gamma}{A} \right. \right. \\ \left. \left. + \frac{1}{qr} \left(\frac{N^2(N-1)}{2|B|} + \frac{i\gamma(N-1)}{A} \right) + \delta_1 ar \left(-\frac{2\gamma}{N+2} \left(\frac{1}{A} - \frac{2}{|B|} \right) + \frac{iN}{|B|} \right) \right\} \right. \\ \left. + \mathfrak{G}_k^* \left\{ \text{same expression with} \right. \right. \\ \left. \left. \text{the sign of } |B| \text{ changed} \right\} \right] \times (N+1) a_0 r^{\beta+1} e^{iqr-r/N a_0} dr, \end{aligned} \quad (36A)$$

and

$$\begin{aligned} \frac{2\pi}{3} \int_0^\infty \left[\mathfrak{G}_{-k-1} \left\{ \gamma(N-1) \left(\frac{1}{A} + \frac{1}{|B|} \right) - \frac{a\delta_1}{q} \frac{N}{|B|} + \frac{iN^2(N-1)}{2} \left(\frac{3}{A} - \frac{1}{|B|} \right) \right. \right. \\ \left. \left. - \frac{ia\delta_1}{q} \frac{2}{N+2} \frac{\gamma}{A} + \frac{1}{qr} \left(\frac{N^2(N-1)}{2|B|} + \frac{i\gamma(N-1)}{A} \right) \right. \right. \\ \left. \left. + \delta_1 ar \left(-\frac{2\gamma}{N+2} \left(\frac{1}{A} + \frac{1}{|B|} \right) - iN \left(\frac{3}{A} - \frac{1}{|B|} \right) \right) \right\} \right. \\ \left. + \mathfrak{G}_{-k-1}^* \left\{ \text{same expression with} \right. \right. \\ \left. \left. \text{the sign of } |B| \text{ changed} \right\} \right] \times (N+1) a_0 r^{\beta+1} e^{iqr-r/N a_0} dr. \end{aligned} \quad (36B)$$

These expressions may be evaluated as before. Expression (36A) contains six integrals

$$X_1 = \delta_1 \int_0^\infty \mathcal{G}_k e^{iqr - r/Na_0} r^{\beta+1} \cdot ar \, dr,$$

$$Y_1 = \delta_1 \int_0^\infty \mathcal{G}_k^* e^{iqr - r/Na_0} r^{\beta+1} \cdot ar \, dr,$$

and four integrals P, Q, R and S, which are the same as P, Q, R and S of § 4 with δ_1 replacing δ (which implies Na_0 replacing a_0). The last four are calculated exactly as before. For a given value of $h\nu$, however, the values of the constants involved are slightly different, since the energy of the emitted electron is different. We have

$$E = h\nu (1 + \frac{1}{2}\theta N)$$

so that

$$\begin{aligned} a^2 &= q^2 \{1 + \theta N + \theta^2 (N^2/4 - 1)\} \\ &= q^2/\tau_1^2, \text{ say,} \end{aligned}$$

and the constants b and c have different values. Remembering this, P and Q, may be calculated from (20) and (21) and R, S obtained from the former by the formulæ (27), provided δ and τ are replaced by δ_1 and τ_1 in both calculations.

The integrals X_1 and Y_1 are found from (18) and (22) respectively, by replacing β by $\beta + 1$ and multiplying by δ_1 . They may be expressed in terms of P_1 and Q_1 by using the same recurrence formulæ (26). We find

$$\left. \begin{aligned} M^* X_1 [\delta_1 - i(\tau_1 - 1)] &= \delta_1 (\beta + 1 - ib) M^* P_1 \\ &\quad + \delta_1 (s + 1 - ib) M Q_1 \\ \text{and} \\ M Y_1 [\delta_1 - i(\tau_1 - 1)] (1 - z) &= \delta_1 (s + 1 + ib) M^* P_1 \\ &\quad + \delta_1 (\beta + 1 + ib) M Q_1. \end{aligned} \right\} \quad (37)$$

When the final state is of type (4B) the work is exactly the same providing we replace s by s' in the above formulæ. These integrals we call X'_1, \dots, S'_1 .

For the final result we introduce as before the quantities $J_1, J'_1, \dots, J_6, J'_6$, defined by

$$P_1 = |(k - s) - i(b + c)| L_1 \cdot J_1, \quad P'_1 = |(b - c) + i(k - 1 - s')| L'_1 J'_1,$$

etc., where L_1 and L'_1 are obtained from L and L' by replacing δ by δ_1 and τ by τ_1 , and the constants b and c are calculated for the L_1 -electrons.

On including the normalisation factors we finally obtain for the internal conversion coefficient the value

$$I_{L_I} = \frac{1}{24} \frac{\gamma \theta b}{Z} (2\gamma)^{2\beta+2} \frac{1}{(\beta+2) \Gamma(2\beta+3)} \frac{N+2}{2(N-1) N^{2\beta+4}} \\ \times \frac{|\Gamma(\beta+1-ib)|^2}{\left| \left\{ \frac{\gamma}{N} + \frac{i}{\theta} \left(\frac{a}{q} - 1 \right) \right\}^{2(\beta+1-ib)} \right|} [2|\mathfrak{A}_1|^2 + |\mathfrak{B}_1|^2]. \quad (38)$$

The expression for \mathfrak{A}_1 and \mathfrak{B}_1 are cumbersome and will not be written down. They are obtained from the expressions inside the square brackets of (36A) and (36B) by replacing \mathfrak{G}_k , \mathfrak{G}_k/qr and $\delta_1 ar \mathfrak{G}_k$ by J_1 , J_3 and J_5 , and \mathfrak{G}_k^* , \mathfrak{G}_k^*/qr and $\delta_1 ar \mathfrak{G}_k^*$ by J_2 , J_4 and J_6 . In (36B), \mathfrak{G}_{-k-1} is replaced by J'_1 , etc.

The asymptotic form of (38) as $\lambda \rightarrow 0$ is easily obtained. As before we find

$$|J_1| \rightarrow 1 \\ J_2, J_3, J_4 \rightarrow 0,$$

and from equations (37)

$$X_1 \rightarrow P_1 \frac{(\beta+1-ib) \delta_1}{[\delta_1 - i(\tau_1 - 1)]} \\ \rightarrow P_1 \frac{(\beta+1-i\gamma)\gamma}{\gamma + i(\beta+2)}$$

so that

$$J_5 \rightarrow J_1 \frac{(\beta+1-i\gamma)\gamma}{\gamma + i(\beta+2)}.$$

Also

$$\frac{\gamma}{N} + \frac{i}{\theta} \left(\frac{a}{q} - 1 \right) \rightarrow \left(\frac{\gamma}{N} + \frac{iN}{2} \right),$$

and on substituting these in (38) we obtain, after some algebra,

$$I_{L_I \text{ Asymptotic}} = \frac{(N+2)}{2(N-1) N^{2\beta+4}} e^{-2\gamma \tan^{-1} [\gamma/(\beta+2)]} \times I_{K \text{ Asymptotic}},$$

both referring to one electron. For radium C this ratio is about 0.149 or 1/6.7.

§ 6. *Internal Conversion by the L_{II} and L_{III} Shells.*—The two L_{II} electrons are given by $n' = 1$, $k = 1$, $u = 0$ and -1 , the wave functions being of type (4B). The selection rule is $\Delta u = 0$ and $\Delta k = \pm 1$, but actually there are only two possible final states for each initial one. They are $k = 0$, type (4A)

and $k = 2$, type (4B). The calculation is similar to that for the L_I shell, but we shall only give the asymptotic formula. We find

$$\frac{I \text{ asymptotic for } L_{II} \text{ shell}}{I \text{ asymptotic for } L_I \text{ shell}} = \frac{(2 - N)(N - 1)}{(2 + N)(N + 1)},$$

where N has the same value as in § 5. The numerical value of this ratio is about 0.0086.

The L_{III} shell consists of four electrons, whose wave functions are of type (4A) with $n' = 0$, $k = 1$ and $u = 0, -1, 1$ and -2 . We have the same selection rule but only three final states are possible for each initial one, since there are no states of type (4B) with $k = 0$. For the four electrons of the L_{III} shell we find asymptotically

$$I_{\text{for } L_{III} \text{ shell}} = \frac{\gamma^2 \theta}{Z} (2\gamma)^2 \sqrt{(4 - \beta^2)} \frac{|\Gamma\{\sqrt{(4 - \beta^2)} - i\gamma\}|^2}{\Gamma\{1 + 2\sqrt{(4 - \gamma^2)}\}} \times \frac{1}{|\{\gamma + i\sqrt{(4 - \gamma^2)}\}^2[\sqrt{(4 - \gamma^2)} - i\gamma]|},$$

which is the same as the formula for the K shell if we replace $\sqrt{(4 - \gamma^2)}$ by $\sqrt{(1 - \gamma^2)}$. The ratio of the internal conversion in the L_{III} shell to that in the L_I shell is about 0.044.

These ratios are only valid in the limit $\lambda \rightarrow 0$ but we should expect them to give the general characteristics of the absorption in the experimental range of wave-lengths. For the ratio I_K to I_{L_I} we find the asymptotic value is very close to that calculated for a definite value of λ , given by $mc^2/h\nu = 0.709$.

Owing to the roughness of the model, we should not expect the values obtained for the L shells to be very accurate. The neglect of the screening effect results in the coincidence of the L_I and L_{II} lines, since they form a "screening doublet."

In the above calculations we have again kept the direction of the nuclear dipole fixed. If we fix the direction of quantisation and then average over all directions of the dipole, the result must be the same. This follows from the fact that the L_{II} and L_{III} shells are spherically symmetrical even though the states of the separate electrons are not.

§ 7. *Results and Discussion.*—The internal conversion coefficient for the K shell has been calculated for six values of $h\nu$, and the results are given in the Table I.

The experimental values are taken from a table given by Ellis and Aston (*loc. cit.*), and the last two refer to radium B. The first four values calculated

Table I.

$\theta = \frac{mc^2}{h\nu}$	$h\nu$ in electron volts $\times 10^{-5}$.	Internal conversion coefficient, calculated.	Internal conversion coefficient, experimental value.
0.287	17.78	0.0008	0.0016
0.452	11.3	0.0018	0.0062
0.639	8.0	0.0035	—
0.835	6.12	0.0057	0.0061
1.443	3.54	0.0172	(0.10)
2.102	2.43	0.0424	(0.25)

are in the region of the radium C lines, and they are shown on a curve, together with the experimental values, in the figure.

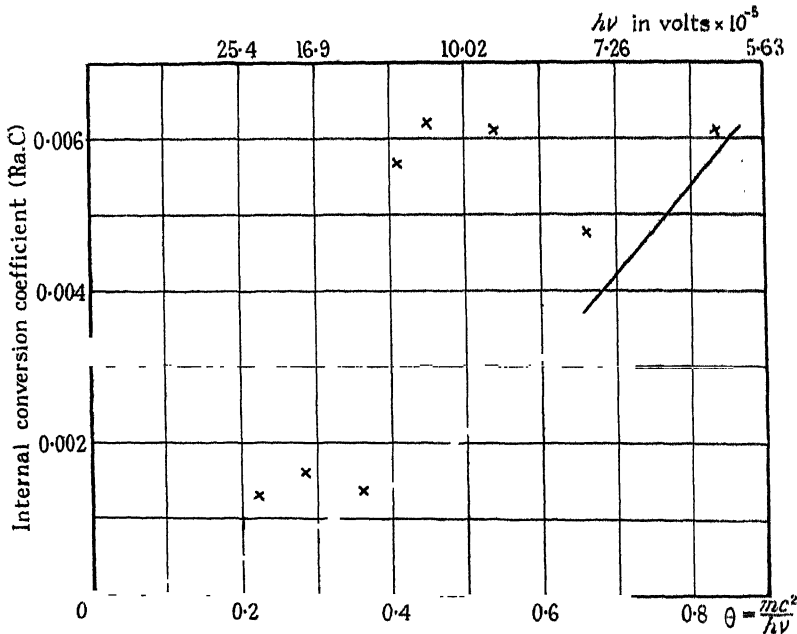


FIG. 1.—The points marked with a cross are the experimental values determined by Ellis and G. H. Aston for radium C for the lines (in volts $\times 10^{-5}$), 6.12, 7.73, 9.41, 11.30, 12.48, 13.89, 17.78 and 22.19.

It will be seen that the curve fits the general order of magnitude if we exclude the three lines between $\theta = 0.4$ and $\theta = 0.6$. Of the remaining lines the three with the lowest energy are certainly within the range of the experimental error.

In the paper following, Taylor and Mott discuss the internal conversion of radiation due to a quadrupole placed at the nucleus. They obtain a curve which passes above the experimental points for the highest energies, and in between the two points for $h\nu = 11.3 \times 10^5$ e.v. and $h\nu = 9.41 \times 10^5$ e.v. It therefore seems probable that the three rays in this region are due to nuclear transitions where the angular momentum of the nucleus changes by two units, and the three rays of lowest energy to transitions where the momentum changes by one unit. For a fuller discussion the reader is referred to the paper following.

It will be seen from the table that the values calculated for radium C are very much smaller than the observed values for radium B. The nuclear charges differ by one unit only, which would reduce the calculated values by about 4 per cent., so that it is quite impossible to explain the internal conversion of radium B on the present dipole theory. Here again Taylor and Mott find that the internal conversion of radiation from a quadrupole is considerably greater in this range, and approaches more nearly the experimental values.

The values of the internal conversion coefficients for the L shells have been calculated for the limit $\lambda \rightarrow 0$. We find $I_K : I_{L_I} : I_{L_{II}} : I_{L_{III}} = 6.7 : 1.0086 : 0.044$, the values being approximate, as we have neglected all screening effects. In the experimental range $I_{L_{III}}$ seems to be always greater than $I_{L_{II}}$ from observed values. It is of great interest to know the value of the ratio I_K, I_{L_I} for wave-lengths in the experimental region. This has been calculated for $\theta = 0.709$, $h\nu = 7.20 \times 10^5$ e.v., and gives a result of about 7.0. As the asymptotic value is 6.7 we are fairly safe in assuming that the ratio does not vary much in the experimental region. Previous theories† gave a value 8 for this ratio, but Dr. Ellis informs me that the present value is in better agreement with experiment.

The results obtained for the absolute magnitude of the internal conversion coefficient differ considerably from those obtained by Miss Swirles and by Casimir. The latter's results are seen to be true in the limit, and some of the terms which vanish as $\lambda \rightarrow 0$ are quite large in the experimental region. One would expect, however, that the non-relativistic results of Miss Swirles would not differ considerably from the present ones in the radium B region. Actually the present theory gives results which are about four times as large, which may seem surprising. We must remember, however, that the relativity correction, which arises from the term $Z/137$, occurs in the wave functions of both initial

† Swirles, *loc. cit.*, and Fowler, *loc. cit.*

and final states. This may be small in the case of light atoms and for the outer electrons of heavier atoms, but for the K shell of radioactive atoms there is no reason why its effect should not be very great.

We may conclude that the internal conversion of γ -rays from the nucleus of radium C is capable of being explained as an ordinary photoelectric effect, provided we exclude the ray with $h\nu = 14.26 \times 10^5$ e.v., which seems to be almost completely converted. In some cases the γ -rays are due to transitions where the angular momentum of the nucleus changes by one unit, while Taylor and Mott have shown that the assumption of quadrupole radiation will account for the remaining rays within the limits of the experimental accuracy.

In conclusion I wish to thank Professor R. H. Fowler for suggesting that I should check Casimir's calculations, and for his interest and encouragement throughout the work, and Dr. C. D. Ellis for opportunities of discussing the experimental values. The numerical work, which is largely due to Mr. J. McDougall, has been calculated by one of us and checked by the other. The algebra for the K shell has been checked by Mr. H. M. Taylor, and I take this opportunity of thanking him and Mr. McDougall for all their assistance. I should also like to thank Mr. Mott and Mr. Taylor for informing me of their results for quadrupole radiation.

Summary.

The internal conversion coefficient for radium C has been calculated on the assumption that the radiation field may be represented by a dipole situated at the nucleus. The accurate relativistic wave functions have been used and the internal conversion coefficient for the K shell has been calculated for several lines. The results are in good agreement with experiment for some lines, which are therefore presumably due to transitions in which the total angular momentum of the nucleus changes by one unit. According to Mott and Taylor the other lines are explainable on the assumption of quadrupole radiation. The ratios of the conversion in the K, L_I , L_{II} and L_{III} shells are calculated for the limit $\lambda \rightarrow 0$, and the ratio $I_K : I_{L_I}$, for a value of $mc^2/h\nu = 0.709$.

A Theory of the Internal Conversion of γ -Rays.

By H. M. TAYLOR, B.A., Clare College, Cambridge, and N. F. MOTT, M.A.,
Gonville and Caius College, Cambridge.

(Communicated by R. H. Fowler, F.R.S.—Received August 20, 1932.)

The "internal conversion coefficient" of a given γ -ray is defined† as the probability that the γ -ray will be absorbed by one of the planetary electrons of the atom. If we denote by α the internal conversion coefficient, and by A the probability per unit time of the emission of a γ -ray by the nucleus (the Einstein A coefficient), then the number of electrons ejected per unit time is $A\alpha$, and the number of quanta escaping unabsorbed is $A(1 - \alpha)$. The quantity actually measured is the ratio of these two, namely

$$\alpha/(1 - \alpha).$$

Experimental values of α have been obtained by Ellis and Aston for eight of the γ -rays of Radium C, and three for Radium B. For Radium C the lines measured lie between 6 and 22×10^5 electron volts; the internal conversion coefficients lie between 0.006 and 0.001, and do *not* vary smoothly with the frequency. For three lines of Radium B of energy in the neighbourhood of 3×10^5 electron volts, α is much bigger, of order of magnitude 0.2.

The problem of internal conversion has been treated theoretically by Miss Swirles,‡ Casimir§ and Hulme.|| These authors assume that the excited nucleus radiates the field of a dipole, namely,

$$\left. \begin{aligned} A_0 &= Br^{-1} e^{2\pi i(r/\lambda - \nu t)} \cos \theta \left(1 + \frac{i}{qr}\right) + \text{complex conjugate} \\ A_z &= Br^{-1} e^{2\pi i(r/\lambda - \nu t)} + \text{complex conjugate} \\ A_x &= A_y = 0, \end{aligned} \right\} \quad (1.1)$$

where

$$q = 2\pi\nu/c.$$

The probability per unit time of the ejection of an electron with energy $h\nu$ is then calculated, using the methods employed in the treatment of the photo-electric effect; the internal conversion coefficient is obtained by dividing by $16\pi^2 B^2 \nu / 3hc$, the number of quanta per unit time radiated by the field (1.1).

† Rutherford, Chadwick and Ellis, "Radiations from Radioactive Substances," p. 510.

‡ 'Proc. Roy. Soc.,' A, vol. 116, p. 491 (1927).

§ 'Nature,' vol. 126, p. 953 (1930).

|| Previous paper, p. 643. Referred to as H II.

The dipole field (1.1) is the field emitted according to quantum mechanics by a radiating system in transitions in which the quantum number j , specifying the total angular momentum†, changes by unity. We may say at once that it is immaterial whether we suppose that the nucleus first emits a quantum, which is then absorbed by a planetary electron, or whether we suppose that the electron is ejected by direct interaction with the nucleus. Whichever way we choose to consider the problem, the probability of ejection of an electron will be given by the method outlined above.‡ Errors may arise, however, from the assumption that the dipole field (1) is correct for all r however small. For points within the nucleus ($r < 10^{-12}$ cm.) this will not be the case; Fowler (*loc. cit.*) has estimated the correction likely to arise, assuming fields of a reasonable magnitude inside the nucleus, and found it to be small for dipole transitions, $\Delta j = \pm 1$. We discuss other transitions below (§ 3).

This correction for $r < 10^{-12}$ cm. has been described by Gamow§ as ejection due to the electron coming right into the nucleus. We would emphasise, however, that it is better to regard it as a correcting term to the internal conversion coefficient, which may either increase or decrease its value.

Miss Swirles and Casimir (*loc. cit.*), using various approximations for the wave functions of the atomic electron, obtain values of α about 10 times too small, but Hulme (*loc. cit.*), using the relativistic equation of Dirac to obtain the wave function of the planetary electron, obtains a value of α which varies smoothly with ν , and gives results in excellent agreement with experiment for the γ -rays of Ra C of energy ($\times 10^5$ e. volts) 6.12, 7.73, 13.89. For the γ -rays of energy ($\times 10^5$ e. volts) 9.41, 11.30, 12.48, and for the rays of Ra B, Hulme's values are too small by a factor of about 6.

The main purpose of this paper is to repeat Hulme's calculations assuming for the field of the nucleus the field of a quadripole, instead of that of a dipole. We confine ourselves to electrons ejected from the K level. The field of a quadripole is

$$\left. \begin{aligned} A_0 &= -Cr^{-1} e^{2\pi i(r/\lambda - \nu t)} \left\{ 2P_2(\cos \theta) \left[1 + \frac{3i}{qr} - \frac{3}{q^2 r^2} \right] + 1 \right\} \\ &\quad + \text{complex conjugate} \\ A_z &= -3Cr^{-1} e^{2\pi i(r/\lambda - \nu t)} \cos \theta \left[1 + \frac{i}{qr} \right] + \text{complex conjugate} \end{aligned} \right\} \quad (1.2)$$

The constant C may have any value.

† j denotes the orbital angular momentum of the radiating particle, assuming only one particle to be excited.

‡ This has been emphasised by Fowler, 'Proc. Roy. Soc.,' A, vol. 129, p. 1 (1930); cf. also § 3 of this paper.

§ 'Atomic Nuclei,' p. 76.

A radiating system emits the field of a quadrupole in transitions where the quantum number j changes by two, or, in certain cases, by zero. For atoms such transitions occur very much less frequently than transitions for which $\Delta j = 1$. It has been pointed out by Gamow and Delbrück, however, that if a nucleus is made up of α -particles only, the electric centre coincides with the centre of gravity, and the dipole moment vanishes. The result of this is, that $\Delta j = 1$ transitions still emit the field (1), and $\Delta j = 2$ transitions emit the field (2), but now the probability per unit time of the $\Delta j = 1$ transition is *less* than that of the $\Delta j = 2$. The probability of $\Delta j = 3$ transitions, etc., is still much smaller than that of the other two. The same result follows from the model, recently assumed for the nucleus by Heisenberg,[†] as is shown in § 4.

We think it reasonable to assume, therefore, that a given γ -ray will have either a quadrupole or a dipole field associated with it, but that octopole, etc., lines do not occur.

Nuclear selection rules are discussed further in § 3, where it is shown that transitions $j = 1 \rightarrow j = 1$ also radiate a quadrupole field. Transitions $j = 0 \rightarrow j = 0$ are absolutely forbidden; this may account for the abnormally large conversion coefficient of the line 14.26×10^5 electron volts, as suggested by Fowler.

The values of the internal conversion coefficient calculated for the quadrupole and dipole[‡] fields are shown in fig. 1.

The curves shown in the two figures are the same, but drawn on a different scale. They are calculated (§ 2) for $Z = 84$; the correction for Ra B ($Z = 83$) should be very small. The experimental values are marked with a cross, Ra B on one figure, Ra C on the other. For θ greater than about 1.8, the quadrupole values are obtained by extrapolation, as shown by the dotted line.

Dr. Ellis has informed us that of the Ra C lines, those of energies 6.12, 11.3, $\times 10^5$ volts are the most reliable. The values of θ for these two lines are 0.84, 0.45. It will be seen that the internal conversion coefficient of the former line lies almost on the dipole curve, and that of the latter on the quadrupole curve. Of the remaining Ra C lines that at 17.78×10^5 volts is fairly reliable; that at 13.89 is less good, and the fact that the published value lies exactly on the theoretical curve must be regarded as fortuitous. One may say that although the experimental results are not yet sufficiently advanced to give a definite proof to the theory, they are certainly consistent with it, and the theory

[†] 'Z. Physik,' vol. 77, p. 1 (1932).

[‡] We are indebted to Mr. H. R. Hulme for permission to publish the dipole curve, calculated by him.

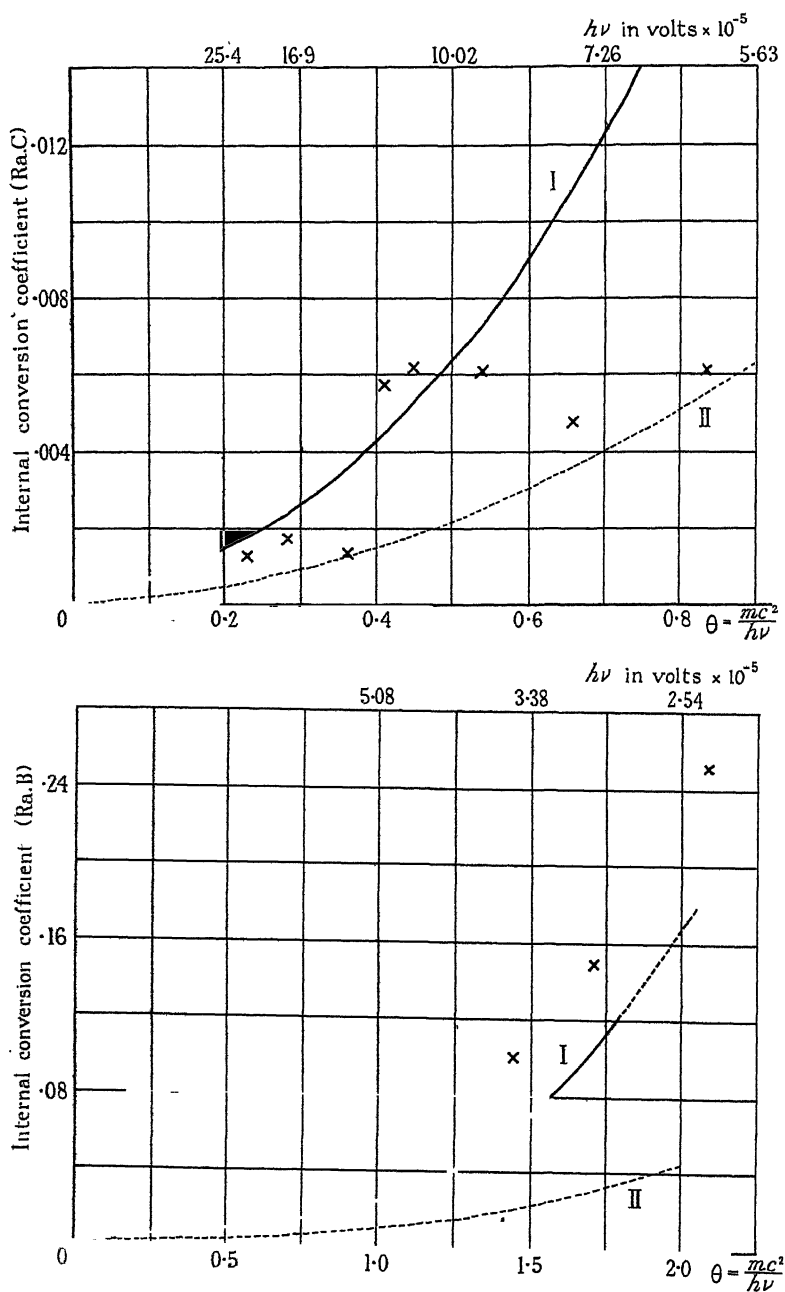


FIG. 1.—The internal conversion coefficient, $\kappa\alpha$ calculated for atomic number $Z = 84$. Curve I is for quadripole radiation, curve II for dipole radiation (calculated by Hulme). The points marked with a cross are the experimental values determined by Ellis and Aston for the lines (in volts $\times 10^{-5}$), 6.12, 7.73, 9.41, 11.30, 12.48, 13.89, 17.78, 22.19 for Ra C, and 2.43, 2.97, 3.54 for Ra B.

does give a reasonable explanation of the fact that the internal conversion coefficient is not a smooth function of the frequency ν .

Dr. Ellis has also informed us that the relative values of α for Ra B are probably more accurate than for Ra C, because the softer γ -rays are more absorbed, and their intensity is therefore easier to measure. The correction for scattering and loss of energy for the photo-electrons emerging from the platinum will, however, be more important for smaller energy, and it is therefore possible that there is a systematic error in the experimental results. It will be seen in fig. 1 that the theory gives results about two-thirds of the experimental, but that the variation with frequency is the same for both.

In conclusion, we think that the agreement between theory and experiment makes it probable that for many of the γ -ray lines, the ejection of electrons is due to ordinary photoelectric absorption; and further, that if this is the case, a nucleus can radiate either the dipole or the quadripole field. This latter result is of considerable importance for nuclear theory; for instance, as we shall show in § 4, in a nucleus built of protons and electrons moving with comparable velocities, transitions giving quadripole radiation will not occur. If on the other hand, the electrons move fast compared to the protons, as assumed by Heisenberg (*loc. cit.*) and as one would expect by analogy with a molecule such as H_2 , the dipole moment is small, and dipole and quadripole transitions may be equally probable.

§ 2. *Calculation of the Internal Conversion Coefficient.*—We calculate the number of electrons ejected per unit time from the K ring of the atom by the method of Variation of Parameters.† For the initial state of the K electron, we take the solution of Dirac's relativistic equation for an electron in the field of a nucleus of charge Ze ; thus the interaction with the other extra-nuclear electrons is neglected. We take for the perturbing field the field of a quadripole; the calculation is otherwise similar to that of Hulme (H II) for the dipole case.

The quadripole field is (§§ 4 and 5)

$$\left. \begin{aligned} A_0 &= \mathcal{A}_0 + \mathcal{A}_0^* & A &= \mathcal{A} + \mathcal{A}^* \\ \text{where} \quad \mathcal{A}_0 &= -Cr^{-1}e^{i(qr-2\pi\nu t)} \left\{ 2P_2(\cos\theta) \left[1 + \frac{3i}{qr} - \frac{3}{q^2r^2} \right] + 1 \right\} \\ \mathcal{A}_x &= \mathcal{A}_y = 0 \\ \mathcal{A}_z &= -3Cr^{-1}e^{i(qr-2\pi\nu t)} \cos\theta \left[1 + \frac{i}{qr} \right] \end{aligned} \right\}, \quad (2.1)$$

† Cf. Dirac, 'Quantum Mechanics,' p. 162, or Gaunt, 'Phil. Trans.,' A, vol. 229, p. 192 (1929).

and the number of quanta emitted per second by such a field is

$$\frac{24\pi}{5} \cdot \frac{q}{h} \cdot C^2, \quad (2.2)$$

where

$$q = 2\pi\nu/c.$$

The perturbing field causes transitions from the initial stationary state of the electron to states in which the electron is free and has an energy E given by

$$E = E_0 + h\nu, \quad (2.3)$$

where E_0 is the energy of the initial state. Now let ψ_i be the wave function (without time factor) of the initial state, normalised to unity, and let ψ_f be the wave function also without time factor for any one such free state, this wave function being so normalised that one electron per unit time crosses an infinite sphere about the atom (Gaunt, *loc. cit.*). Then the probability p per unit time of transitions between these two states, due to the perturbing potentials (2.1), is given by

$$p = \left(\frac{2\pi}{h}\right)^2 |(\psi_f^* | \epsilon A_0 + \epsilon \rho_1(A, \sigma) | \psi_i)|^2, \quad (2.4)$$

where ρ_1 and σ are the matrices of four rows introduced by Dirac† and where the time factors may now be suppressed in A_0 and A .

In order to obtain the total probability of emission of electrons, we must first sum (2.4) over all possible final states of the electron, and then multiply by two on account of the two electrons of the K ring. Thus the probability, $\kappa\alpha$, per quantum of radiation from the nucleus,‡ of the ejection of a β -particle, is given by

$$\kappa\alpha = \sum_f 2 \cdot \frac{5}{24\pi} \cdot \frac{h}{q} \cdot \frac{1}{C^2} \left(\frac{2\pi\epsilon}{h}\right)^2 |(\psi_f^* | A_0 + \rho_1(A, \sigma) | \psi_i)|^2. \quad (2.5)$$

Furthermore, of the two constituent parts \mathcal{A} and \mathcal{A}^* of A only the first is effective in causing transitions of the electron in which its energy changes from E_0 to $E_0 + h\nu$, the second being effective in transitions to states of energy $E_0 - h\nu$. Hence, for our purpose, we may substitute in formula (2.5), for A and A_0 , the values \mathcal{A} and \mathcal{A}_0 given in (2.1).

We now proceed to show what final states are possible for the electron, and

† Dirac, 'Quantum Mechanics,' p. 242.

‡ This is the definition of $\kappa\alpha$ as given in Rutherford, Chadwick and Ellis, p. 510.

to calculate the corresponding values of the matrix elements in (2.5). Convenient explicit forms for the wave functions, both of the free and the bound states, have been given by Hulme,[†] and we shall follow his notation here. Of the two possible initial states for the electron (H I equations (44) and (45)) we take for the present the first only, corresponding to an electron with its spin along the positive z -axis, and discuss later the question of an arbitrary combination of the two states. Hence we put

$$\left. \begin{aligned} (\psi_i)_1 &= -if \cos \theta & (\psi_i)_2 &= -if \sin \theta e^{i\phi} \\ (\psi_i)_3 &= g & (\psi_i)_4 &= 0 \end{aligned} \right\} \quad (2.6)$$

where

$$f = \frac{\gamma}{1 + (1 - \gamma^2)^{\frac{1}{2}}} \xi(E_0) r^\beta e^{-r/a_0}$$

and

$$g = \xi(E_0) r^\beta e^{-r/a_0},$$

$\xi(E_0)$ being the normalising factor, γ being $2\pi e^2 Z/hc$ and β being $(1 - \gamma^2)^{\frac{1}{2}} - 1$.

Since $\mathcal{S}_x = \mathcal{S}_y = 0$, (A, σ) reduces simply to $\mathcal{S}_z \sigma_z$, and therefore in order to evaluate the matrix element, we require only the values of ψ_f^* , ψ_i and $\psi_f^* \rho_1 \sigma_z \psi_i$ both these being summed over all four components. Now there are two possible types for the final state (H I, equations (3A) and (3B)), given by

$$\left. \begin{aligned} (\psi_f)_1 &= -iF_k P_k^u & (\psi_f)_2 &= -iF_k P_{k+1}^{u+1} \\ (\psi_f)_3 &= (k+u+1) G_k P_k^u & (\psi_f)_4 &= (-k+u) G_k P_k^{u+1} \end{aligned} \right\} \quad (2.7A)$$

and

$$\left. \begin{aligned} (\psi_f)_1 &= -i(k+u) F_{-k-1} P_{-k-1}^u & (\psi_f)_2 &= -i(-k+u+1) F_{-k-1} P_{-k-1}^{u+1} \\ (\psi_f)_3 &= G_{-k-1} P_k^u & (\psi_f)_4 &= G_{-k-1} P_k^{u+1} \end{aligned} \right\}, \quad (2.7B)$$

and we divide the discussion into two cases, according as the final state is of type A or type B. We denote by $\kappa\alpha_A$ the total probability of transitions to final states of the type A and by $\kappa\alpha_B$ the probability of transitions to states of the type B, so that

$$\kappa\alpha = \kappa\alpha_A + \kappa\alpha_B. \quad (2.8)$$

Since the functions P_k^u in the above equations are the Legendre functions *including the term* in $e^{iu\phi}$ as introduced by Darwin[‡], it is at once clear that, for final states of *either* type, the matrix element (2.4) vanishes on account of the integration with respect to ϕ unless u has the value zero.

[†] Hulme, 'Proc. Roy. Soc.,' A, vol. 133, p. 381 (1931), referred to as H I.

[‡] Darwin, 'Proc. Roy. Soc.,' A, vol. 118, p. 668 (1928).

Case A.—Final state of the type (2.7A) with $u = 0$. Let us introduce the function \mathcal{P}_k^u defined by

$$\mathcal{P}_k^u = (k-u)! \sin^u \theta \left(\frac{d}{d \cos \theta} \right)^{k+u} \frac{(\cos^2 \theta - 1)^k}{2^k k!}, \quad (2.9)$$

from which it follows that $\mathcal{P}_k^u e^{iu\phi} = P_k^u$. Then we have the following results:

$$\left. \begin{aligned} \psi_f^* \psi_i &= fF_k \{ \cos \theta \mathcal{P}_{k+1}^0 + \sin \theta \mathcal{P}_{k+1}^1 \} + (k+1) gG_k \mathcal{P}_k^0 \\ \psi_f^* \rho_1 \sigma_z \psi_i &= i \{ gF_k \mathcal{P}_{k+1}^0 - fG_k [(k+1) \cos \theta \mathcal{P}_k^0 + k \sin \theta \mathcal{P}_k^1] \} \end{aligned} \right\} \quad (2.10)$$

From this it follows that

$$\int_0^\pi \mathcal{A}_0 \psi_f^* \psi_i \sin \theta d\theta \quad \text{and} \quad \int_0^\pi \mathcal{A}_z \psi_f^* \rho_1 \sigma_z \psi_i \sin \theta d\theta$$

are zero† unless $k = 2$. Hence there is only one possible final state of type A for the ejected electron, namely, that determined by $u = 0$, $k = 2$. Putting $k = 2$ in equations (2.10), we find, after some transformation

$$\left. \begin{aligned} \psi_f^* \psi_i &= (fF_2 + gG_2) \cdot 6P_2(\cos \theta) \\ \psi_f^* \rho_1 \sigma_z \psi_i &= i \{ 6P_3(\cos \theta) (gF_2 - \frac{1}{3}fG_2) - \frac{24}{5}fG_2P_1(\cos \theta) \} \end{aligned} \right\} \quad (2.11)$$

in which $P_n(\cos \theta)$ is the ordinary Legendre function. Inserting these values and the values of the potentials from (2.1) in (2.4) and performing the integration with respect to θ , we obtain

$$\begin{aligned} &(\psi_f^* | A_0 + \rho_1(A, \sigma) | \psi_i) \\ &= \frac{48\pi C}{5} \int_0^\infty \left\{ - (fF_2 + gG_2) \left(1 + \frac{3i}{qr} - \frac{3}{q^2 r^2} \right) + 2ifG_2 \left(1 + \frac{i}{qr} \right) \right\} e^{i\alpha r} r dr. \end{aligned} \quad (2.12)$$

Now Hulme has shown‡ (H I equation (8)) that

$$\left. \begin{aligned} \text{and} \quad &AF_k + i|B|G_k = \mathcal{G}_k \\ &AF_k - i|B|G_k = \mathcal{G}_k^* \end{aligned} \right\} \quad (2.13)$$

† The case $k = 0$ calls for special attention. The integrals do not appear to vanish, but if we use the identity $\text{div } j + 1/c \cdot \partial \rho / \partial t = 0$ they may be transformed so as to vanish identically. If the perturbing potentials had been taken in the unmodified form of (4.15), (4.19) and (4.20) then the integrals for $k = 0$ would vanish without further transformation.

‡ We refer to H I for the definitions of A , B and \mathcal{G} .

and if we substitute from these relations in (2.12) and also insert the values of f and g from (2.6) we obtain

$$(\psi^*_f | A_0 + \rho_1(A, \sigma) | \psi_i) = \frac{24\pi C_\xi^2(E_0)}{5(\beta + 2)} \left\{ A |B| \sqrt{\left(\frac{2}{A^2 + |B|^2} \right)} \right\}^{-1} \mathfrak{B}_2, \quad (2.14)$$

where

$$\begin{aligned} \mathfrak{B}_2 = & \int_0^\infty \{ [(2\gamma + i \cdot \overline{\beta + 2}) D - \gamma E] \mathfrak{G}_2 \\ & + [(-2\gamma - i \cdot \overline{\beta + 2}) D - \gamma E] \mathfrak{G}_2^* \} e^{iqr - r/a_0} r^{\beta+1} dr \\ & + \int_0^\infty \frac{1}{qr} \{ [(-3 \cdot \overline{\beta + 2} + 2i\gamma) D - 3ie \dots \gamma E] \mathfrak{G}_2 \\ & + [(3 \cdot \overline{\beta + 2} - 2i\gamma) D - 3i\gamma E] \mathfrak{G}_2^* \} e^{iqr - r/a_0} r^{\beta+1} dr \\ & + \int_0^\infty \frac{1}{q^2 r^2} \{ 3(-i \cdot \overline{\beta + 2} D + \gamma E) \mathfrak{G}_2 \\ & + 3(i \cdot \overline{\beta + 2} D + \gamma E) \mathfrak{G}_2^* \} e^{iqr - r/a_0} r^{\beta+1} dr, \end{aligned} \quad (2.15)$$

in which we have put

$$\left. \begin{aligned} D &= A \left(\frac{2}{A^2 + |B|^2} \right)^{\frac{1}{2}} \\ E &= |B| \left(\frac{2}{A^2 + |B|^2} \right)^{\frac{1}{2}} \end{aligned} \right\} \quad (2.16)$$

Now let us write

$$\left. \begin{aligned} \int_0^\infty \mathfrak{G}_k e^{iqr - r/a_0} r^{\beta+1} dr &= P_k \\ \int_0^\infty \frac{1}{qr} \mathfrak{G}_k e^{iqr - r/a_0} r^{\beta+1} dr &= R_k \\ \int_0^\infty \frac{1}{q^2 r^2} \mathfrak{G}_k e^{iqr - r/a_0} r^{\beta+1} dr &= T_k \end{aligned} \right\}, \quad (2.17)$$

and let us denote by Q_k , S_k and U_k the corresponding integrals with \mathfrak{G}_k^* substituted for \mathfrak{G}_k ; let us further denote by \mathcal{P}_2 , \mathcal{Q}_2 , \mathcal{R}_2 , \mathcal{S}_2 , \mathcal{T}_2 and \mathcal{U}_2 the coefficients of P_2 , Q_2 , etc., in (2.15), so that we have

$$\mathcal{P}_2 = \{2\gamma + i(\beta + 2)\} D - \gamma E, \quad \text{etc.}, \quad (2.18)$$

and

$$\mathfrak{B}_2 = \mathcal{P}_2 P_2 + \mathcal{Q}_2 Q_2 + \dots + \mathcal{U}_2 U_2. \quad (2.19)$$

Then Hulme has shown (H II (15), (20), (21) and (23)) how to transform P_2 and Q_2 into hypergeometric series whose values we may determine by numerical

computation, and has shown (H II, (27))† that R_2 and S_2 may then be determined from the relations

$$\left. \begin{aligned} \frac{q}{a} (s + \beta + 1) (s - \beta + 1) \frac{z}{2i} R_2 &= (-\beta - ib) P_2 \\ &+ (s + 1 - ib) (1 - z) \frac{M}{M^*} Q_2 \\ \text{and} \\ \frac{q}{a} (s + \beta + 1) (s - \beta + 1) \frac{z}{2i} S_2 &= (s + 1 + ib) \frac{M^*}{M} P_2 \\ &+ (-\beta + ib) (1 - z) Q_2 \end{aligned} \right\} \quad (2.20)$$

By writing $\beta - 1$ for β we may at once obtain similar relations giving the values of T_2 and U_2 in terms of R_2 and S_2 , namely,

$$\left. \begin{aligned} \frac{q}{a} (s + \beta) (s - \beta + 2) \frac{z}{2i} T_2 &= (-\beta - ib + 1) R_2 \\ &+ (s + 1 - ib) (1 - z) \frac{M}{M^*} S_2 \\ \text{and} \\ \frac{q}{a} (s + \beta) (s - \beta + 2) \frac{z}{2i} U_2 &= (s + 1 + ib) \frac{M^*}{M} R_2 \\ &+ (-\beta + 1 + ib) (1 - z) S_2 \end{aligned} \right\} \quad (2.21)$$

All the constants occurring in the equations (2.15)–(2.20) are determined in terms of the wave-length ($2\pi/q$) of the nuclear γ -ray, and, therefore, by means of these equations we may calculate the value of β_2 corresponding to any given wave-length of γ -ray.

We have still to multiply the matrix element (2.14) by the normalising factor $\xi(E, k, u)$ for the final state of the electron. Substituting the matrix element, so normalised,‡ in (2.5), we obtain

$$\kappa \alpha_A = \frac{48\pi h \xi^2(E_0) \xi^2(E, k, u)}{5q(\beta + 2)^2} \left(\frac{2\pi\epsilon}{h} \right)^2 \frac{A^2 + |B|^2}{2A^2 |B|^2} |\beta_2|^2.$$

Case B.—Final state of the type (2.7B) with $u = 0$. For this case we have the following results

$$\left. \begin{aligned} \psi_f^* \psi_i &= f F_{-k-1} \{k \cos \theta \mathcal{P}_{k-1}^0 + (1 - k) \sin \theta \mathcal{P}_{k-1}^1\} + g G_{-k-1} \mathcal{P}_k^0 \\ \text{and} \\ \psi_f^* \rho_1 \sigma_z \psi_i &= i \{k g F_{-k-1} \mathcal{P}_{k-1}^0 + f G_{-k-1} (-\cos \theta \mathcal{P}_k^0 + \sin \theta \mathcal{P}_k^1)\} \end{aligned} \right\} \quad (2.22)$$

† We again refer to H II for the definitions of a, b, s, z, M and its conjugate complex M^* . We may note here that $2\pi/a$ is the wave-length of the wave associated with the free electron while, of course, $2\pi/q$ is the wave-length of the γ -ray.

‡ The values of $\xi(E_0)$ and $\xi(E, k, u)$ are given in (H I), equations (46) and (22A).

from which it again follows as in Case A that the transition probability is zero except for $k = 2$. Giving k this value, and performing the integration over the angle θ , we obtain, in this case,

$$(\psi^*_r | A_0 + \rho_1(A, \sigma) | \psi_i) = -\frac{8\pi C}{5} \int_0^\infty \left\{ 2(fF_{-3} + gG_{-3}) \left(1 + \frac{3i}{qr} - \frac{3}{q^2 r^2} \right) + i(5gF_{-3} + fG_{-3}) \left(1 + \frac{i}{qr} \right) \right\} e^{iar} r dr. \quad (2.23)$$

Making the substitutions (2.13), (2.16) and (2.17) we obtain this matrix element in the form

$$-\frac{4\pi C \xi(E_0)}{5(\beta + 2)} \left\{ A | B | \sqrt{\left(\frac{2}{A^2 + |B|^2} \right)} \right\}^{-1} \mathcal{B}_{-3}, \quad (2.24)$$

where $\mathcal{B}_{-3} = \mathcal{P}_{-3} P_{-3} + \dots + \mathcal{U}_{-3} U_{-3}$, and where the values of the coefficients are given by

$$\left. \begin{aligned} \mathcal{P}_{-3} &= \{\gamma - 2i(\beta + 2)\} D + \{2\gamma + 5i(\beta + 2)\} E \\ \mathcal{Q}_{-3} &= \{-\gamma + 2i(\beta + 2)\} D + \{2\gamma + 5i(\beta + 2)\} E \\ \mathcal{R}_{-3} &= \{6(\beta + 2) + i\gamma\} D + \{-5(\beta + 2) + 6i\gamma\} E \\ \mathcal{S}_{-3} &= -\{6(\beta + 2) + i\gamma\} D + \{-5(\beta + 2) + 6i\gamma\} E \\ \mathcal{T}_{-3} &= +6i(\beta + 2) D & -6\gamma E \\ \mathcal{U}_{-3} &= -6i(\beta + 2) D & -6\gamma E \end{aligned} \right\}. \quad (2.25)$$

The relations (2.20) and (2.21) giving R, S, T and U in terms of P and Q remain unaltered except that for s we must now write s' (see H I equations (17A) and (17B)). Hence these equations, together with (2.24) and (2.25) provide the means of calculating the value of \mathcal{B}_{-3} corresponding to any given γ -ray. The corresponding contribution to the internal conversion coefficient is given by

$$\kappa_{\alpha B} = \frac{4\pi\hbar \xi^2(E_0) \xi^2(E, -k-1, u)}{15q(\beta + 2)^2} \left(\frac{2\pi\epsilon}{\hbar} \right)^2 \frac{A^2 + |B|^2}{2A^2 |B|^2} |\mathcal{B}_{-3}|^2. \quad (2.26)$$

We have now to discuss the question of the other possible initial state. Both initial states are of the type (2.7A), and the one we have already considered is given by $k = 0, u = 0$. The alternative state, given by $k = 0$ and $u = -1$ corresponds to the electronic spin being directed along the negative z -axis, and has the wave function

$$\left. \begin{aligned} (\psi_i)_1 &= if \sin \theta e^{-i\phi} & (\psi_i)_2 &= -if \cos \theta \\ (\psi_i)_3 &= 0 & (\psi_i)_4 &= -g \end{aligned} \right\}. \quad (2.27)$$

Let us denote by Ψ_0 and Ψ_{-1} these two initial states (2.6) and (2.27). Then from (2.27) it follows that, if the electron is initially in the state Ψ_{-1} , the matrix elements are zero for all final states except such as are specified by $u = -1$ and $k = 2$. On working out the matrix elements for these transitions we find that they have the same values as before, (2.12) and (2.23), and therefore the transition probabilities κ_{A_0} and κ_{B_0} also have the same values. That is to say that if $\psi_{A,0}$, $\psi_{A,-1}$, $\psi_{B,0}$ and $\psi_{B,-1}$ denote final states of the types A and B, with $k = 2$ in all cases and $u = 0$ or -1 , respectively, and if V denotes the perturbation $A_0 + \rho_1(A, \sigma)$ then

$$\left. \begin{aligned} (\psi_{A,0}^* | V | \Psi_0) &= (\psi_{A,-1}^* | V | \Psi_{-1}) = P_A, \\ (\psi_{B,0}^* | V | \Psi_0) &= (\psi_{B,-1}^* | V | \Psi_{-1}) = P_B \end{aligned} \right\} \quad (2.28)$$

and the matrix elements vanish for all other arrangements of the final states.

Now an electron with its spin in an arbitrary direction is described by the wave function $(A\Psi_0 + B\Psi_{-1})/N$, where A and B are complex constants depending on the direction of the spin, and N is the normalising factor, which, since Ψ_0 and Ψ_{-1} are already normalised and orthogonal, has the value $\{|A|^2 + |B|^2\}^{\frac{1}{2}}$. Hence the matrix elements contributing to κ_{A_0} arising from such an initial state are

$$N^{-2} \sum_u (\psi_{A,u} | V | A\Psi_0 + B\Psi_{-1})^2, \quad (2.29)$$

which by (2.28) yields

$$N^{-2} (|A|^2 + |B|^2) P_A = P_A. \quad (2.30)$$

In similar fashion

$$N^{-2} \sum_u (\psi_{B,u} | V | A\Psi_0 + B\Psi_{-1})^2 = P_B, \quad (2.31)$$

and so the total transition probability is seen to be independent of the initial orientation of the electronic spin. There is no need now to perform any averaging over directions of orientation of the quadripole field (2.1), for we may take the z -axis, or axis of quantisation of the electronic system, to coincide with the axis of the field since we have shown that the orientation of electronic spin relative to the axis of quantisation is immaterial.

The Physical Significance of the Two Possible Final States.—We may use the vector model† to form a convenient picture of what is happening to the electron in the two cases A and B. We denote by \mathbf{s} , \mathbf{l} and \mathbf{j} the angular momentum vectors corresponding to the spin, the orbital momentum and their resultant,

† Pauling and Goudsmid, "Structure of Line Spectra," pp. 32 and 33. We follow their notation.

and by s , l and j the magnitudes of these vectors. The resolved parts along the z -axis we denote by \mathbf{m}_s , \mathbf{m}_l and \mathbf{m} . Then Darwin† has pointed out that j and $|\mathbf{m}|$ are respectively the averages of subscripts and superscripts of the spherical harmonics in the equations (2.7A) and (2.7B). Further, by considering the passage to the non-relativistic limit, it is clear that l is in each case the subscript of the harmonic occurring in ψ_3 and ψ_4 .

For the initial state (2.6), with $k = 0$ and $u = 0$ we have therefore $\mathbf{m} = \frac{1}{2}$, $j = \frac{1}{2}$ and $l = 0$ corresponding to an electron with no orbital angular momentum and with its spin along the positive z -axis. The corresponding final state A is described by $\mathbf{m} = \frac{1}{2}$, $j = 5/2$ and $l = 2$ while the final state B is described by $\mathbf{m} = \frac{1}{2}$, $j = 3/2$ and $l = 2$. In fig. 2 we show the combinations of vectors which these quantum numbers imply, and we see that, while in each case the z -component of total angular momentum remains unchanged, the two cases differ in that for A the spin remains along the positive z -axis, while for B it is changed to point in the opposite direction.

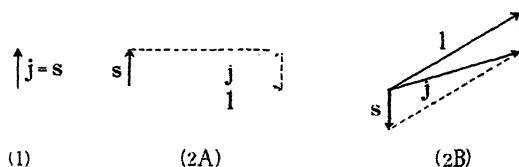


FIG. 2.—(1) Initial state, $\mathbf{m}_s = \frac{1}{2}$, $\mathbf{m}_l = 0$, $\mathbf{m} = \frac{1}{2}$, $j = \frac{1}{2}$, $l = 0$. (2A) Final state $\mathbf{m}_s = \frac{1}{2}$, $\mathbf{m}_l = 0$, $\mathbf{m} = \frac{1}{2}$, $j = 5/2$, $l = 2$. (2B) Final state, $\mathbf{m}_s = -\frac{1}{2}$, $\mathbf{m}_l = 1$, $\mathbf{m} = \frac{1}{2}$, $j = 3/2$, $l = 2$.

We should therefore expect the probability for transitions to the final state B to be much smaller than the probability for transitions to the state A, and this is, in fact, the case, except for very short wave-length, as is shown in fig. 3, where the two probabilities are shown separately. The internal conversion coefficient of fig. 1 is, of course, obtained by adding the ordinates of points on these two curves of fig. 3.

We must add that for very short wave-lengths, $h\nu \gg mc^2$, the formula for internal conversion with the quadripole field tends to the same value as that obtained with the dipole field by Casimir (*loc. cit.*), as may be verified by the method of H II, § 4.

§ 3. The purpose of this section is to discuss the theoretical basis upon which the calculation of the preceding sections rests, and to show that it is immaterial whether we consider that the electron is ejected by direct interaction with the

† *Loc. cit.*, p. 666.

nucleus or by absorption of a γ -ray. We also discuss the corrections which may arise due to the finite size of the nucleus.

For purposes of discussion we take a model for the nucleus, assuming the

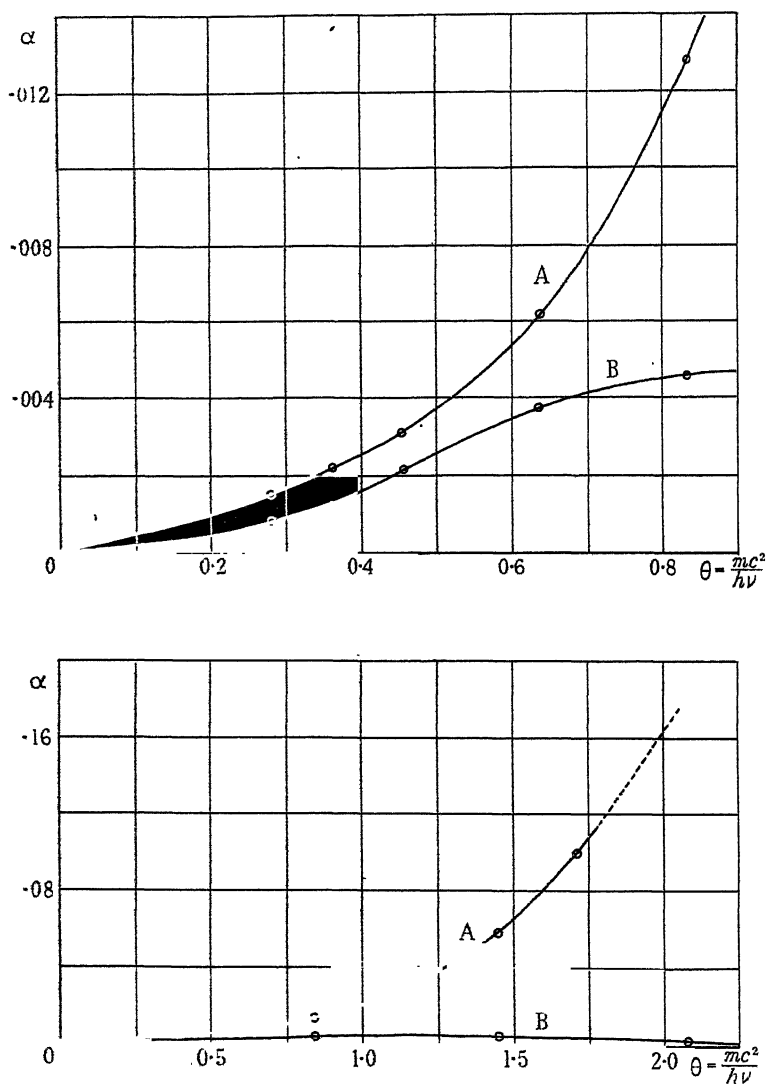


FIG. 3.—Curve A, $\kappa\alpha_A$; curve B, $\kappa\alpha_B$.

radiation to be emitted by a single α -particle moving in a fixed field. We denote, by \mathbf{R} the co-ordinate of the α -particle, and by $\chi_i(\mathbf{R})$ the wave function of its initial, excited state, and by $\chi_f(\mathbf{R})$ the wave function of its final state after emitting the γ -ray. We denote by \mathbf{r} the co-ordinate of the K electron, by

$\psi_i(\mathbf{r})$ the wave function, normalised to unity, of its initial state in the K ring, and by $\psi_f(\mathbf{r})$ the wave function of the final free state, normalised so that one electron crosses a large sphere in either direction per unit time.

Assuming that the α -particle is a point charge, the interaction between the α -particle and the electron is given by

$$V(\mathbf{r}, \mathbf{R}) = -2e^2/|\mathbf{r} - \mathbf{R}|.$$

According to non-relativistic quantum mechanics, the probability per unit time that the nucleus will make the transition $i \rightarrow f$, thereby ejecting an electron, is

$$\frac{4\pi^2}{\hbar^2} \left| \iint d\mathbf{r} d\mathbf{R} \chi_{*f}^*(\mathbf{R}) \psi_{*f}^*(\mathbf{r}) V(\mathbf{r}, \mathbf{R}) \chi_i(\mathbf{R}) \psi_i(\mathbf{r}) \right|^2. \quad (3.1)$$

The calculation is the familiar one of the transition between states of equal energy. No mention is made of the emission of any light quantum.

Möller† has shown, in his paper on the collision between two electrons, how to introduce retarded potentials into the problem. In our case we must write down the charge density ρ and current density \mathbf{j} associated with the nuclear transition $i \rightarrow f$. These are

$$\begin{aligned} \rho &= e\chi_{*f}^*\chi_i(\mathbf{R}) \exp[2\pi i(W_f - W_i)t/\hbar] + \text{complex conjugate,} \\ \text{and} \\ \mathbf{j} &= \frac{e\hbar}{4\pi m c i} (\chi_{*f}^* \text{grad } \chi_i - \chi_i \text{grad } \chi_{*f}^*) \exp[2\pi i(W_f - W_i)t/\hbar] \\ &\quad + \text{complex conjugate.} \end{aligned}$$

We then find the scalar potential A_0 and vector potential \mathbf{A} due to this charge and current. A_0 and \mathbf{A} satisfy

$$\left. \begin{aligned} \nabla^2 A_0 - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} A_0 &= 4\pi\rho \\ \nabla^2 \mathbf{A} - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} \mathbf{A} &= 4\pi\mathbf{j} \end{aligned} \right\}, \quad (3.2)$$

but (3.2) are not sufficient to determine A_0 , \mathbf{A} . We impose the further condition that A_0 , \mathbf{A} must represent an outgoing wave; we then obtain

$$\begin{aligned} A_0 &= e^{-2\pi i\nu t} e \int \frac{\exp iq|\mathbf{R} - \mathbf{r}|}{|\mathbf{R} - \mathbf{r}|} \chi_{*f}^*(\mathbf{R}) \chi_i(\mathbf{R}) d\mathbf{R} \\ \hbar\nu &= W_i - W_f, \quad q = 2\pi\nu/c, \end{aligned} \quad (3.3)$$

† 'Z. Physik,' vol. 70, p. 786 (1932).

with a similar expression for A . We then calculate the effect of the periodic field (3.3) on the electron. The calculation is the same as that used for the photoelectric effect; the number of electrons ejected per unit time is

$$\left(\frac{2\pi}{h}\right)^2 \left| \int (\psi_{*f} | -\varepsilon A_0 - \varepsilon \rho_1 A \cdot \sigma | \psi_i) d\tau \right|^2. \quad (3.4)$$

We note that if we make $c \rightarrow \infty$, A vanishes and q is small, so that (3.4) tends to (3.1).

The field (3.3) may be considered as the field radiated by the atom; in this sense one may say that a quantum is given out and reabsorbed by the electron. However, the integral in (3.4) includes an integration over the space within the nucleus; one cannot separate the part due to direct interaction and the part due to absorption of a quantum.

We now turn to the consideration of the field (3.3) due to the radiating nucleus. If we assume $\chi_i(\mathbf{R})$ to be a wave function of the form $\phi(\mathbf{R}) P_1(\cos \theta)$, the wave function for a P state, and for $\chi_f(\mathbf{R})$ a spherically symmetrical form, then, for points r outside the nucleus (i.e. such that $\chi(\mathbf{R})$ vanishes), (3.3) gives the field (1.1) due to a vibrating dipole. Similarly if we assume for χ_i the wave function for a D state ($j=2$) of the form $\psi(\mathbf{R}) P_2(\cos \theta)$, we obtain the field (1.2) of a quadripole. This is shown in § 4, where the other possible transitions are discussed, such as $P \rightarrow P$. It is further shown in § 5 that the dipole, quadripole forms for the field are the most general outgoing waves possible, which will cause a change of angular momentum 1 and 2 respectively in the K electron.

In the calculations of § 2, which are compared with experiment in fig. 1, and also in Hulme's calculations, the field (3.3) is taken to be that of a quadripole or dipole for all r down to $r=0$. They should, therefore, be corrected by a term due to the failure of this field for r less than the radius of the nucleus. We cannot calculate this term, as we do not know the structure of the nucleus. We may, however, estimate its order of magnitude, as has been done by Fowler† for dipole transitions.

The correcting term will depend on the values of the initial and final values of the wave function of the K electron. Since the wave function of an electron with angular momentum j has a zero at the origin of order r^j , the correcting term will be greatest for electronic transitions for which $\Delta j = 0$. Assuming an interaction term of order mc^2 between the nuclear radiator and the electron,

† *Loc. cit.*

it is easy to see that the correcting term will be negligible for any transitions for which Δj is not zero. For nuclear transitions, in which j is zero in the initial and final states, no escape of radiation is possible, and the ejection of electrons must be entirely due to the region within the nucleus.

§ 4. Our purpose in this section is to examine the field radiated by a quantum mechanical system (in our case the nucleus) when it makes transitions in which Δj has various values, and in particular to show that when $\Delta j = 1$ the radiated field is that of a dipole, and for $\Delta j = 2$ or 0 the field is that of a quadrupole. We also examine what happens when the system is built of particles having all the same charge and mass, so that the dipole moment vanishes.

We shall make the following assumptions about the nucleus; firstly, that the radiating particle in the nucleus may be described by a wave function $\psi_{n,l,m}$ which, as for atomic wave functions, may be resolved into radial and angular parts so that

$$\psi_{n,l,m} = \phi_l^n(r) P_l^{|m|}(\cos \theta) e^{im\phi}, \quad (4.1)$$

and secondly, that the nucleus has a finite radius a_N , that is to say, that $\phi(r) = 0$ for all values of r greater than a_N ($a_N \sim 10^{-12}$ cm.).

With these assumptions we derive the electromagnetic radiation from the nucleus corresponding to transitions between D and S states, P and P states, as well as the dipole transition between P and S states. To economise space we shall discuss only the axially symmetrical P and D states, but the analysis for the other P and D states follows along similar lines, and the electromagnetic potentials so derived, although in many cases different in form from the ones we obtain, can be transformed to give the same results. The potentials we derive are valid for all points outside the nucleus, i.e., for $r > a_N$. In § 3 we discussed the correction which should be made to allow for the fact that in § 2 we have assumed these potentials to hold down to $r = 0$.

For brevity, let us describe two stationary states of the nucleus by the two single suffixes 0 and a , assuming that 0 specifies the ground state, and a applies to an excited state. Then the wave functions, with time factors, for these two states are

$$\left. \begin{aligned} u_0 &= \psi_0 e^{-2\pi i E_0 t / \hbar} \\ u_a &= \psi_a e^{-2\pi i E_a t / \hbar} \end{aligned} \right\}, \quad (4.2)$$

where E_0 and E_a are the energies of the two states and where, of course, $E_a > E_0$. Thus the electric density in e.s.u., to be associated with a transition from the excited state to the ground state, is given by

$$\rho = (\psi_a \psi_0^* e^{-2\pi i \nu t} + \psi_a^* \psi_0 e^{2\pi i \nu t}) \varepsilon, \quad (4.3)$$

where $\nu = (E_a - E_0)/\hbar$ and e is the charge of the radiating particle. The current density, in e.m.u., to be associated with the same transition, is given by

$$\mathbf{j} = \frac{e\hbar}{4\pi mc} \left\{ (\psi_0^* \text{grad } \psi_a - \psi_a \text{grad } \psi_0^*) e^{-2\pi i \nu t} / i + (\psi_0 \text{grad } \psi_a^* - \psi_a^* \text{grad } \psi_0) e^{2\pi i \nu t} / (-i) \right\}. \quad (4.4)$$

If now A and A_0 are the vector and scalar potentials of the electromagnetic field, due to the nuclear transition, then A and A_0 satisfy the equations (3.2). We require the solution periodic in the time with period ν . For this solution

$$\left. \begin{aligned} \left\{ \nabla^2 + \left(\frac{2\pi\nu}{c} \right)^2 \right\} A_0 &= -4\pi\rho \\ \left\{ \nabla^2 + \left(\frac{2\pi\nu}{c} \right)^2 \right\} A &= -4\pi\mathbf{j} \end{aligned} \right\}. \quad (4.5)$$

In order to avoid the continual repetition of the conjugate complex constituents of ρ , \mathbf{j} , A_0 and A , let us write

$$\left. \begin{aligned} A_0 &= \mathcal{A}_0 + \mathcal{A}_0^* \\ A &= \mathcal{A} + \mathcal{A}^* \end{aligned} \right\}, \quad (4.6)$$

and let us also write $q = 2\pi\nu/c = 2\pi(E_a - E_0)/\hbar c$. Then, substituting in (4.5) from (4.3) and (4.4) and using Green's Theorem, we find that the solutions of (4.5), which represent waves travelling out from the nucleus, are given by

$$\left. \begin{aligned} \mathcal{A}_0(\mathbf{r}) &= e \int \frac{e^{iq|\mathbf{r}-\mathbf{r}_1|}}{|\mathbf{r}-\mathbf{r}_1|} \psi_a(\mathbf{r}_1) \psi_0^*(\mathbf{r}_1) e^{-2\pi i \nu t} d\tau_1 \\ \text{and} \\ \mathcal{A}(\mathbf{r}) &= \frac{e\hbar}{4\pi mc} \int \frac{1}{i} \frac{e^{iq|\mathbf{r}-\mathbf{r}_1|}}{|\mathbf{r}-\mathbf{r}_1|} \left\{ \psi_0^*(\mathbf{r}_1) \text{grad}_1 \psi_a(\mathbf{r}_1) \right. \\ &\quad \left. - \psi_a(\mathbf{r}_1) \text{grad}_1 \psi_0^*(\mathbf{r}_1) \right\} e^{-2\pi i \nu t} d\tau_1 \end{aligned} \right\}, \quad (4.7)$$

where \mathbf{r} is the position vector of the point at which the potentials are being measured, \mathbf{r}_1 is the position vector of a point in the nucleus, grad_1 denotes differentiation with respect to the co-ordinates of this point, and where the volume integration is to be extended throughout the nucleus. That is to say, that the volume integration is to be taken from $r_1 = 0$ to $r_1 = a_N$ since the ψ 's vanish for $r_1 > a_N$.

In order to transform these two expressions conveniently, we shall make use of two lemmas which may be easily proved.

Lemma I.—If for values of r greater than a_N the function $F(r, \theta, \phi)$ is defined by the relation

$$F(r, \theta, \phi) = \int_0^{a_N} \int_0^\pi \int_0^{2\pi} \frac{e^{iq|\mathbf{r}-\mathbf{r}_1|}}{|\mathbf{r}-\mathbf{r}_1|} f(r_1) P_l^{|m|}(\cos \theta_1) e^{im\phi_1} r_1^2 dr_1 \sin \theta_1 d\theta_1 d\phi_1,$$

then, for $r > a_N$,

$$F(r, \theta, \phi) = 2i\pi^2 \frac{H_{l+1/2}(qr)}{\sqrt{r}} P_l^{|m|}(\cos \theta) e^{im\phi} \times \int_0^{a_N} \frac{J_{l+1/2}(qr_1)}{\sqrt{r_1}} f(r_1) r_1^2 dr_1, \quad (4.8)$$

where $H_{l+1/2}$ is the "first Hankel function" defined by the relation†

$$H_{l+1/2}(z) = J_{l+1/2}(z) - i(-)^l J_{-l-1/2}(z).$$

It is convenient to give here, for purposes of reference, the values of the first three Hankel functions of half-integral order:—

$$\left. \begin{aligned} H_{1/2}(z) &= \frac{1}{i} \left(\frac{2}{\pi z} \right)^{1/2} e^{iz} \\ H_{3/2}(z) &= - \left(\frac{2}{\pi z} \right)^{1/2} e^{iz} \left\{ 1 + \frac{i}{z} \right\} \\ H_{5/2}(z) &= i \left(\frac{2}{\pi z} \right)^{1/2} e^{iz} \left\{ 1 + \frac{3i}{z} - \frac{3}{z^2} \right\} \end{aligned} \right\} \quad (4.9)$$

Lemma II.—If $\phi_l(r)$ and $\phi_0(r)$ are solutions, which vanish for $r \geq a_N$, of the equations

$$\left. \begin{aligned} \phi_l''(r) + \frac{2}{r} \phi_l'(r) + \left\{ \frac{8\pi^2 m}{\hbar^2} [E_a - V(r)] - \frac{l(l+1)}{r^2} \right\} \phi_l(r) &= 0 \\ \text{and} \\ \phi_0''(r) + \frac{2}{r} \phi_0'(r) + \frac{8\pi^2 m}{\hbar^2} [E_0 - V(r)] \phi_0(r) &= 0 \end{aligned} \right\} \quad (4.10)$$

and if $q = 2\pi(E_a - E_0)/\hbar c$, then

$$\int_0^{a_N} r^j (\phi_l' \phi_0 - \phi_l \phi_0') r^2 dr = \frac{1}{j+1} \left\{ \frac{4\pi m c q}{\hbar} \int_0^{a_N} r^{j+1} \phi_l \phi_0 r^2 dr - l(l+1) \int_0^{a_N} r^{j-1} \phi_l \phi_0 r^2 dr \right\}. \quad (4.11)$$

We shall now develop in detail the electromagnetic potentials for the parti-

† Cf. Watson, "Theory of Bessel Functions," p. 73.

cular nuclear transitions in which the initial state is a D state determined by the wave function†

$$\psi_a = \psi_{n, l=2, m=0} = \phi_2(r) P_2^0(\cos \theta), \quad (4.12)$$

and the ground state is the normal S state,

$$\psi_0 = \psi_{n=0, l=0, m=0} = \phi_0(r). \quad (4.13)$$

From (4.7) we have, for this transition,

$$\mathcal{A}_0(\mathbf{r}) = \varepsilon e^{-2\pi i \mathbf{r} \cdot \mathbf{t}} \int_0^{a_N} \int_0^\pi \int_0^{2\pi} \frac{e^{i\mathbf{q} \cdot (\mathbf{r} - \mathbf{r}_1)}}{|\mathbf{r} - \mathbf{r}_1|} \Phi_2(r_1) \phi_0(r_1) P_2(\cos \theta) r_1^2 dr_1 \sin \theta_1 d\theta_1 d\phi_1,$$

which, by Lemma I, yields

$$\mathcal{A}_0(\mathbf{r}) = 2i\pi^2 \varepsilon e^{-2\pi i \mathbf{r} \cdot \mathbf{t}} r^{-1/2} H_{5/2}(qr) P_2(\cos \theta) \int_0^{a_N} r_1^{-1/2} J_{5/2}(qr_1) \phi_2(r_1) \phi_0(r_1) r_1^2 dr_1. \quad (4.14)$$

Now even for the hardest γ -rays from radium C, q is less than 10^{11} cm.^{-1} , and if we assign to the nucleus a radius of about 10^{-12} cm. ,‡ then qr is less than one-tenth for all points inside the nucleus. Hence we may expand the Bessel function under the integral sign in (4.14) in powers of qr , and the leading term in $\mathcal{A}_0(r, \theta, \phi)$ will come from the lowest power of qr .

Substituting the value of $H_{5/2}$ from (4.9) and expanding $J_{5/2}$ as

$$(2/\pi)^{1/2} \cdot z^{5/2}/15 \cdot \{1 - z^2/14 + \dots\},$$

we obtain

$$\begin{aligned} \mathcal{A}_0(r, \theta, \phi) = & -\frac{4\pi}{15} r^{-1} e^{i(qr-2\pi \mathbf{r} \cdot \mathbf{t})} \left[1 + \frac{3i}{qr} - \frac{3}{q^2 r^2} \right] P_2(\cos \theta) \\ & \times \int_0^{a_N} q^2 r^2 \left(1 - \frac{q^2 r^2}{14} + \dots \right) \phi_0 \phi_2 r^2 dr, \end{aligned} \quad (4.15)$$

and therefore the leading term in the scalar potential is

$$\mathcal{A}_0(r, \theta, \phi) = -2C_2 P_2(\cos \theta) r^{-1} e^{i(qr-2\pi \mathbf{r} \cdot \mathbf{t})} \left[1 + \frac{3i}{qr} - \frac{3}{q^2 r^2} \right], \quad (4.16)$$

where

$$C_2 = \frac{2\pi}{15} \int_0^{a_N} \varepsilon q^2 r^2 \phi_0(r) \phi_2(r) r^2 dr.$$

The vector-potential is most simply discussed by considering its Cartesian

† Where there is no danger of confusion we drop the upper suffix to the ϕ symbol. Thus in (4.12) we should, in the notation of (4.1) write ϕ_2^n but for brevity we may drop the n .

‡ Cf. Gamow, "Atomic Nuclei and Radioactivity," p. 52.

components. For this special transition we have from (4.6) that the z -component of the vector-potential is determined by

$$\mathcal{A}_z(\mathbf{r}) = \frac{\varepsilon \hbar e^{-2\pi i \nu t}}{4\pi m c i} \int_0^{a_N} \int_0^\pi \int_0^{2\pi} \frac{e^{iq|\mathbf{r}-\mathbf{r}_1|}}{|\mathbf{r}-\mathbf{r}_1|} \left\{ \phi_0(r_1) \frac{\partial}{\partial z_1} [\phi_2(r_1) P_2(\cos \theta_1) - \phi_2(r_1) P_2(\cos \theta_1)] \frac{\partial}{\partial z_1} \phi_0(r_1) \right\} r_1^2 dr_1 \sin \theta_1 d\theta_1 d\phi_1. \quad (4.17)$$

The partial derivatives of functions such as occur in (4.17) have been given in a convenient form by Darwin,[†] and using his results, the quantity inside the curly bracket of (4.17) may be transformed to yield

$$\frac{2}{3} P_3 \left\{ \phi_0 \phi'_2 - \frac{2\phi_0 \phi_2}{r} - \phi'_0 \phi_2 \right\} + \frac{2}{3} P_1 \left\{ \phi_0 \phi'_2 + \frac{3\phi_0 \phi_2}{r} - \phi'_0 \phi_2 \right\} \quad (4.18)$$

where the dashes denote differentiation with respect to r . Substituting this quantity in (4.17) and applying Lemma I we obtain

$$\mathcal{A}_z(r, \theta, \phi) = \frac{\varepsilon \hbar}{4\pi m c} \cdot 2\pi^2 \{ \lambda(r) P_3(\cos \theta) + \mu(r) P_1(\cos \theta) \} e^{-2\pi i \nu t},$$

where

$$\lambda(r) = \frac{2}{3} r^{-1/2} H_{7/2}(qr) \int_0^{a_N} r^{-1/2} J_{7/2}(qr) \left\{ \phi_0 \phi'_2 - \frac{2\phi_0 \phi_2}{r} - \phi'_0 \phi_2 \right\} r^2 dr$$

and

$$\mu(r) = \frac{2}{3} r^{-1/2} H_{3/2}(qr) \int_0^{a_N} r^{-1/2} J_{3/2}(qr) \left\{ \phi_0 \phi'_2 + \frac{3\phi_0 \phi_2}{r} - \phi'_0 \phi_2 \right\} r^2 dr. \quad (4.19)$$

If now we expand the Bessel functions, as in the previous case, we at once see that the integral in $\lambda(r)$ is small compared with that in $\mu(r)$ in the ratio of $q^2 a^2$ to 1, and therefore we may reject the term in $\lambda(r)$. Further, we notice that the leading term of $\mu(r)$, namely,

$$\int_0^{a_N} qr \left\{ \phi_0 \phi'_2 + \frac{3\phi_0 \phi_2}{r} - \phi'_0 \phi_2 \right\} r^2 dr$$

may be transformed by Lemma II to yield

$$\frac{1}{2} \cdot \frac{4\pi m c}{\hbar} \int_0^{a_N} q^2 r^2 \phi_0 \phi_2 r^2 dr$$

[†] Darwin, 'Proc. Roy. Soc.,' A, vol. 118, p. 668 (1928). We use here the third of his equations (7.2). It might be noted that we are here using ordinary Legendre functions, not the special ones as defined by Darwin.

and, therefore, finally we have

$$\mathcal{A}_z(r, \theta, \phi) = -2 C_2 P_1(\cos \theta) r^{-1} e^{i(qr-2\pi\nu t)} \left[1 + \frac{i}{qr} \right], \quad (4.20)$$

where C_2 is as defined in (4.16).

In a precisely similar fashion, using the first of Darwin's equations (7.2) (*loc. cit.*), we may transform the corresponding expression for $\mathcal{A}_{\omega+iy}$, and in this way we obtain

$$\left. \begin{aligned} \mathcal{A}_x &= C_2 P_1^1(\cos \theta) \cos \phi \cdot r^{-1} e^{i(qr-2\pi\nu t)} \left[1 + \frac{i}{qr} \right] \\ \mathcal{A}_y &= C_2 P_1^1(\cos \theta) \sin \phi \cdot r^{-1} e^{i(qr-2\pi\nu t)} \left[1 + \frac{i}{qr} \right] \end{aligned} \right\}. \quad (4.21)$$

The electromagnetic potentials corresponding to a given field of radiation are, of course, not unique. As is well known, if we modify our potentials by substituting

$$\left. \begin{aligned} \mathcal{A}'_0 &= \mathcal{A}_0 + \frac{1}{c} \frac{\partial \lambda}{\partial t} \\ \mathcal{A}' &= \mathcal{A} - \text{grad } \lambda \end{aligned} \right\} \quad (4.22)$$

where λ is any scalar, satisfying

$$\nabla^2 \lambda - \frac{1}{c^2} \frac{\partial^2 \lambda}{\partial t^2} = 0,$$

then the potentials A' and A'_0 defined in terms of \mathcal{A}' and \mathcal{A}'_0 by (4.6) still lead to the same values for the field intensities E and H . Let us choose λ to satisfy

$$\lambda = C_2 / iqr \cdot e^{i(qr-2\pi\nu t)}. \quad (4.23)$$

Substituting from (4.16), (4.20) and (4.21) in (4.22) we therefore obtain

$$\left. \begin{aligned} \mathcal{A}'_0 &= -C_2 r^{-1} e^{i(qr-2\pi\nu t)} \left\{ 2 P_2(\cos \theta) \left[1 + \frac{3i}{qr} - \frac{3}{q^2 r^2} \right] + 1 \right\} \\ \text{and} \quad (\mathcal{A}')_x &= (\mathcal{A}')_y = 0 \\ (\mathcal{A}')_z &= -3C_2 r^{-1} e^{i(qr-2\pi\nu t)} \cos \theta \left[1 + \frac{i}{qr} \right] \end{aligned} \right\} \quad (4.24)$$

where

$$C_2 = \frac{2\pi}{15} \int_0^{\infty} \varepsilon q^2 r^2 \phi_0 \phi_2 r^2 dr.$$

This is the ordinary field of a quadrupole. Cf. § 5.

Comparison with other Transitions.—In a method, precisely similar to that developed above, we may derive the potentials radiated by any transition of the nucleus. We shall write down here the potentials corresponding to a transition from a P state to the ground state, and also those corresponding to a transition between two P states, whose radial wave functions we shall write in the notation of (4.1) as ϕ_1^a and ϕ_1^b .

P \rightarrow S transition.

$$\left. \begin{aligned} \mathcal{A}_0 &= -C_1 r^{-1} e^{i(qr-2\pi\nu t)} \cos \theta \\ \mathcal{A}_x &= \mathcal{A}_y = 0 \\ \mathcal{A}_z &= -C_1 r^{-1} e^{i(qr-2\pi\nu t)} \end{aligned} \right\} \quad (4.25)$$

where

$$C_1 = \frac{4\pi}{3} \int_0^{a_N} \varepsilon q r \phi_0 \phi_1^a r^2 dr.$$

P \rightarrow P transition.

$$\left. \begin{aligned} \mathcal{A}_0 &= -C_{11} r^{-1} e^{i(qr-2\pi\nu t)} \left\{ 2P_2(\cos \theta) \left[1 + \frac{3i}{qr} - \frac{3}{q^2 r^2} \right] + 1 \right\} \\ \mathcal{A}_x &= \mathcal{A}_y = 0 \\ \mathcal{A}_z &= -3C_{11} r^{-1} e^{i(qr-2\pi\nu t)} \cos \theta \left[1 + \frac{i}{qr} \right] \end{aligned} \right\} \quad (4.26)$$

where

$$C_{11} = \frac{4\pi}{45} \int_0^{a_N} \varepsilon q^2 r^2 \phi_1^a \phi_1^b r^2 dr.$$

For completeness it should be added that although we have considered in detail only P \rightarrow P and D \rightarrow S transitions, it is true, in general, that for any transitions in which $\Delta j = 0$ or ± 2 the radiation is of the quadrupole type. There is one case which is exceptional, namely, that of a transition between two S states. In this case it is found that by a suitable transformation of the type (4.22) the potentials of the radiation may be made to vanish identically. Hence no *radiative*† transition is possible between two S states.

It will be noticed that the potentials (4.25) are the ordinary potentials corresponding to a Hertzian vibrator along the z -axis. This gives a wave-mechanical justification for the assumption of the field of a Hertzian vibrator made by Miss Swirles, Hulme and Casimir (*loc. cit.*), since this particular field is here shown to correspond to nuclear transitions in which the angular momentum of the nucleus changes by one quantum.

† This point may be of importance in connection with the abnormally converted line at 14.26×10^5 electron volts, as is mentioned in § 1.

The potentials (4.24), corresponding to $D \rightarrow S$ transitions, are seen to be identical with (4.26) except for the unimportant difference in the constants C_2 and C_{11} . These potentials differ, however, from the dipole (4.25) in two important ways which we must now discuss. The first obvious point of difference is the occurrence of second order spherical harmonics in the PP and DS potentials, while in the PS potentials only first order harmonics occur. The presence of the second order harmonics is a characteristic of *quadrupole radiation* just as the first order harmonics are characteristic of *dipole radiation*. In discussing the effect of the fields in causing transitions of extra-nuclear electrons from bound states it is seen that the dipole fields cause transitions in which the orbital angular momentum of the electron changes by one quantum unit whereas the quadrupole fields cause transitions in which the angular momentum of the electron changes by two units (*cf.* Hulme, H II, and § 2 of this paper). The second point of difference between the quadrupole and dipole potentials is the difference in the numerical constants in equations (4.25) and (4.24) or (4.26). We have already pointed out that ga is a small quantity, and therefore C_2 or C_{11} will be small in comparison with C_1 . This corresponds to the fact, well known in considering radiation from atoms, that a dipole transition is normally much more probable than a quadrupole.

As stated in § 1, however, with certain models of the nucleus the dipole moment vanishes. As we shall show below, this has the result that DS transitions ($\Delta j = 2$) become *more* frequent than PS transitions ($\Delta j = 1$.) We must emphasise here, however, that $\Delta j = 1$ transitions will still occur and *will be associated with a dipole field*, as before, so that the calculations of Hulme of the internal conversion coefficient are not affected. The vanishing of the dipole moment affects only the absolute intensities of the lines.

If a radiating system contains particles $\mathbf{R}_1, \mathbf{R}_2, \dots$, with charges $\epsilon_1, \epsilon_2, \dots$, etc., and if in a given transition the initial wave function is ψ_a , the final ψ_0 , the dipole moment is

$$\int [\epsilon_1 \mathbf{R}_1 + \epsilon_2 \mathbf{R}_2 + \dots] \psi_a(\mathbf{R}_1, \mathbf{R}_2, \dots) \psi_0^*(\mathbf{R}_1, \mathbf{R}_2, \dots) d\mathbf{R}_1 d\mathbf{R}_2, \dots$$

This will clearly vanish if ϵ/m is the same for each particle, provided that there is no external field.

It has recently been suggested by Heisenberg† that the nucleus may be built up of protons and neutrons, the forces between them being of the nature of “Austausch” forces, due to the rapid interchange of electrons from neutron

† *Loc. cit.*

to proton and back. It might be thought, at first sight, that with such a model the dipole moment would not vanish; however, closer consideration shows that, in fact, the dipole moment does vanish. The same is true of the hydrogen molecule; no transition in which the electronic state does not change has a dipole moment.

The proof is as follows. If $\mathbf{R}_1, \mathbf{R}_2, \dots$, are the co-ordinates of the protons in a nucleus, $\mathbf{r}_1, \mathbf{r}_2, \dots$, of the electrons, then the wave function of the nucleus may be assumed to be of the form, as for the hydrogen molecule

$$\psi_n(\mathbf{R}_1, \mathbf{R}_2, \dots) \Psi(\mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{r}_1, \mathbf{r}_2, \dots),$$

where ψ is the wave function of the protons moving under the "Austausch" forces, and Ψ the wave function of the electrons moving in the field of the protons, supposed at rest at the points $\mathbf{R}_1, \mathbf{R}_2, \dots$. We may suppose that the function Ψ does not alter in a nuclear transition, and that

$$\int |\Psi|^2 d\mathbf{r}_1 d\mathbf{r}_2 \dots = 1 \quad (\text{all } \mathbf{R}).$$

The dipole moment of the $n \rightarrow 0$ transition will be

$$D = e \iint \{\Sigma \mathbf{R} - \Sigma \mathbf{r}\} \psi_n \psi_0^* |\Psi|^2 d\mathbf{r} d\mathbf{R}. \quad (4.27)$$

But if λ is the ratio of the mass of the proton to the mass of an electron, we have, since the centre of gravity does not move,

$$0 = \iint \{\lambda \Sigma \mathbf{R} + \Sigma \mathbf{r}\} \psi_n \psi_0^* |\Psi|^2 d\mathbf{r} d\mathbf{R}. \quad (4.28)$$

Adding, we obtain

$$D = e(1 + \lambda) \iint \{\Sigma \mathbf{R}\} \psi_n \psi_0^* |\Psi|^2 d\mathbf{r} d\mathbf{R}. \quad (4.29)$$

The integration over \mathbf{r} may be carried out, giving for D

$$\begin{aligned} D &= e(1 + \lambda) \iint \{\Sigma \mathbf{R}\} \psi_n \psi_0^* d\mathbf{R} \\ &= 0. \end{aligned} \quad (4.30)$$

We must now consider in greater detail the radiation from a transition in which $\Delta j = 1$, when the dipole moment vanishes. For this purpose we take a simple model, supposing the nucleus to consist of two interacting particles of mass m_1 and m_2 and charge e_1 and e_2 . These may be thought of as an α -particle, and the rest of the nucleus.

Let the co-ordinates of the two particles relative to a fixed set of axes be (x_1, y_1, z_1) and (x_2, y_2, z_2) , the corresponding position-vectors being \mathbf{r}_1 and \mathbf{r}_2 , and let the co-ordinates of the centre of gravity be (ξ, η, ζ) , the position-vector being \mathbf{R} . Further, let us write

$$\text{so that } \left. \begin{aligned} \mathbf{r} &= \mathbf{r}_1 - \mathbf{r}_2 \\ x &= x_1 - x_2, \text{ etc.} \end{aligned} \right\}, \quad (4.31)$$

and finally let (ξ_1, η_1, ζ_1) and (ξ_2, η_2, ζ_2) be the co-ordinates of the particles relative to axes parallel to the original ones but having their origin at the centre of gravity, so that

$$\left. \begin{aligned} \xi_1 &= m_2 (m_1 + m_2)^{-1} x, \text{ etc.} \\ \xi_2 &= -m_1 (m_1 + m_2)^{-1} x, \text{ etc.} \end{aligned} \right\}. \quad (4.32)$$

Then the wave function $\psi(\mathbf{r}_1, \mathbf{r}_2)$ for the two particles may be resolved into the product of the two functions which describe the motion of the centre of gravity and the motion relative to it; we have

$$\psi(\mathbf{r}_1, \mathbf{r}_2) = \phi(\mathbf{R}) \chi(\mathbf{r}) = \phi(\xi, \eta, \zeta) \chi(x, y, z), \quad (4.33)$$

where $\chi(\mathbf{r})$ satisfies the equation

$$\nabla_r^2 \chi + \frac{8\pi^2 m_1 m_2}{h^2 (m_1 + m_2)} (E_r - V) \chi = 0, \quad (4.34)$$

in which E_r denotes the part of the energy of the whole system which arises from motions relative to the centre of gravity.

Now the charge associated with the particle 1 in the element of co-ordinate space $d\xi d\eta d\zeta dx dy dz$ is

$$\varepsilon_1 |\phi(\xi, \eta, \zeta)|^2 |\chi(x, y, z)|^2 d\xi \dots dz, \quad (4.35)$$

and so the charge in the element of space $dx dy dz$, independent of the position of the centre of gravity, is

$$\begin{aligned} \varepsilon_1 |\chi(x, y, z)|^2 dx dy dz \iiint |\phi(\xi, \eta, \zeta)|^2 d\xi d\eta d\zeta \\ = \varepsilon_1 |\chi(x, y, z)|^2 dx dy dz, \end{aligned} \quad (4.36)$$

since we may suppose ϕ and χ independently normalised to unity. Now let us take the centre of gravity as the origin of co-ordinates, and, therefore, substituting from (4.32) in the expression (4.36) we find that the charge associated with the element $d\xi_1 d\eta_1 d\zeta_1$ arising from the particle 1 is

$$\varepsilon_1 |\chi(\lambda_1 \xi_1, \lambda_1 \eta_1, \lambda_1 \zeta_1)|^2 \lambda_1^3 d\xi_1 d\eta_1 d\zeta_1, \quad (4.37)$$

where $\lambda_1 = (m_1 + m_2)/m_2$, and similarly that the charge in the element $d\xi_2 d\eta_2 d\zeta_2$, due to the particle 2, is

$$\varepsilon_2 |\chi(-\lambda_2 \xi_2, -\lambda_2 \eta_2, -\lambda_2 \zeta_2)|^2 \lambda_2^3 d\xi_2 d\eta_2 d\zeta_2, \quad (4.38).$$

where $\lambda_2 = (m_1 + m_2)/m_1$. If then we denote by \mathbf{u} the position-vector of any point in "centre of gravity space," the total volume density of charge at \mathbf{u} is given by

$$\rho(\mathbf{u}) = \varepsilon_1 \lambda_1^3 |\chi(\lambda_1 \mathbf{u})|^2 + \varepsilon_2 \lambda_2^3 |\chi(-\lambda_2 \mathbf{u})|^2. \quad (4.39)$$

But the wave function $\chi(\mathbf{r})$, being the solution of (4.34) is expressible as in (4.1) in the form

$$\chi_{n,l,m}(\mathbf{r}) = \phi_l^n(r) P_l^{|m|}(\cos \theta) e^{im\phi}, \quad (4.40)$$

and therefore

$$\begin{aligned} \chi_{n,l,m}(-\mathbf{r}) &= \phi_l^n(r) P_l^{|m|}(\cos[\pi - \theta]) e^{im(\phi + \pi)} \\ &= (-)^l \phi_l^n(r) P_l^{|m|}(\cos \theta) e^{im\phi}. \end{aligned} \quad (4.41)$$

Further, since the ground-state χ_0 is spherically symmetrical, we have

$$\chi_0(\mathbf{r}) = \chi_0(-\mathbf{r}).$$

If now we consider transitions in which the system jumps from the state $\chi_{n,l,m}$ to the ground state, then the charge-density at the point \mathbf{u} is given by

$$\begin{aligned} \rho(\mathbf{u}) &= e^{-2\pi i \nu t} \{ \varepsilon_1 \lambda_1^3 \phi_0(\lambda_1 \mathbf{u}) \phi_l^n(\lambda_1 \mathbf{u}) \\ &\quad + (-)^l \varepsilon_2 \lambda_2^3 \phi_0(\lambda_2 \mathbf{u}) \phi_l^n(\lambda_2 \mathbf{u}) \} P_l^{|m|}(\cos \theta) e^{im\phi} \\ &\quad + \text{complex conjugate.} \end{aligned} \quad (4.42)$$

Substituting this expression for ρ in (4.5) and evaluating by means of Lemma I the resulting expression for \mathcal{S}_0 , we obtain

$$\begin{aligned} \mathcal{S}_0(\mathbf{u}) &= 2i\pi^2 e^{-2\pi i \nu t} H_{l+1/2}(qu) P_l^{|m|}(\cos \theta) e^{im\phi} \\ &\quad \times \left\{ \int_0^{a_N/\lambda_1} \varepsilon_1 \lambda_1^3 \phi_0(\lambda_1 r) \phi_l(\lambda_1 r) J_{l+1/2}(qr) r^{3/2} dr \right. \\ &\quad \left. + (-)^l \int_0^{a_N/\lambda_2} \varepsilon_2 \lambda_2^3 \phi_0(\lambda_2 r) \phi_l(\lambda_2 r) J_{l+1/2}(qr) r^{3/2} dr \right\}. \end{aligned} \quad (4.43)$$

The expression inside the curly bracket may be written in the more convenient form

$$\int_0^{a_N} \left\{ \varepsilon_1 \lambda_1^{1/2} J_{l+1/2}(qr/\lambda_1) + (-)^l \varepsilon_2 \lambda_2^{1/2} J_{l+1/2}(qr/\lambda_2) \right\} \phi_0(r) \phi_l(r) r^{3/2} dr, \quad (4.44)$$

and therefore, taking the first term in the expansions of the Bessel functions we obtain the leading term in \mathcal{A}_0 in the form

$$\begin{aligned} \mathcal{A}_0(\mathbf{u}) = & 2i\pi^2 e^{-2\pi i u t} H_{l+1/2}(qu) u^{-1/2} P_l^{|m|}(\cos \theta) e^{im\phi} \\ & \times \frac{(2q/\pi)^{1/2}}{1 \cdot 3 \cdot 5 \dots (2l+1)} (\epsilon_1 \lambda_1^{-l} + (-)^l \epsilon_2 \lambda_2^{-l}) \int_0^{a_N} q^l r^l \phi_0(r) \phi_l(r) r^2 dr, \end{aligned} \quad (4.45)$$

which is identical in form with the expression (4.14) except that the factor $(\epsilon_1 \lambda_1^{-l} + (-)^l \epsilon_2 \lambda_2^{-l})$ replaces the simple quantity ϵ .

Thus we see that when we take into account the motions of the two separate particles, we must multiply the simple expressions previously obtained for the potentials by this factor. For dipole transitions, $l=1$ and the factor becomes $\epsilon_1/\lambda_1 - \epsilon_2/\lambda_2$, while for quadrupole transitions $l=2$ and the factor is $\epsilon_1/\lambda_1^2 + \epsilon_2/\lambda_2^2$.

Since $\lambda_1 : m_1 = \lambda_2 : m_2$ we see that if $\epsilon_1/m_1 = \epsilon_2/m_2$, the leading term in the dipole field vanishes, whereas the order of magnitude of the quadrupole is unaffected. In the dipole case we must therefore take into account the second term in the expansion (4.44). This gives us for $l=1$

$$\begin{aligned} \mathcal{A}_0(\mathbf{u}) = & -2i\pi^2 e^{-2\pi i u t} H_{3/2}(qu) u^{-1/2} P_1^0(\cos \theta) \frac{(2q/\pi)^{1/2}}{30} \left\{ \frac{\epsilon_1}{\lambda_1^3} - \frac{\epsilon_2}{\lambda_2^3} \right\} \\ & \int_0^{a_N} q^3 r^3 \phi_0 \phi_1 r^2 dr, \end{aligned} \quad (4.46)$$

which gives much *less* energy radiated per unit time than in the $l=2$ transitions.

Of course, it is unlikely that the dipole moment vanishes exactly, and it may well be that the energy radiated by $l=1$ transitions is of the same order of magnitude as for $l=2$ transitions, which must be the case if our hypothesis of § 1 as to the various γ -transitions is correct.

§ 5. In this section we discuss the most general form that an electromagnetic wave diverging from the nucleus can have, and show that the assumption of dipole, quadrupole, etc., does not depend on the assumption of any special model for the nucleus.

The electromagnetic field outside the nucleus is conveniently expressed in terms of the "Hertzian Vector," which we denote by Π . The vector-potential \mathbf{A} and the scalar potential A_0 are then given by

$$\mathbf{A} = \frac{1}{c} \frac{\partial}{\partial t} \Pi \quad A_0 = \text{div } \Pi.$$

The advantage of using the Hertzian Vector is that the identity

$$\operatorname{div} \mathbf{A} - \frac{1}{c} \frac{\partial A_0}{\partial t} = 0,$$

is satisfied identically.

The most general solution of Maxwell's equations, representing an outgoing wave of frequency ν , is given by†

$$\Pi = \sum_{\alpha\beta\gamma} C_{\alpha\beta\gamma} \Pi_{\alpha\beta\gamma},$$

where $C_{\alpha\beta\gamma}$ are arbitrary vectors, and

$$\Pi_{\alpha\beta\gamma} = \left(\frac{\partial}{\partial x}\right)^\alpha \left(\frac{\partial}{\partial y}\right)^\beta \left(\frac{\partial}{\partial z}\right)^\gamma \{r^{-1} e^{iqr-2\pi i\nu t}\} + \text{complex conjugate}.$$

It is easily verified that the dipole field (1.1) is given by

$$\Pi_x = \Pi_y = 0, \quad \Pi_z = e^{iqr-2\pi i\nu t}/iqr + \text{complex conjugate},$$

and the quadripole field (1.2) by

$$\Pi_x = \Pi_y = 0, \quad \Pi_z = \frac{\partial}{\partial z} (e^{iqr-2\pi i\nu t}/iqr).$$

It is easily verified (as in H II) that the dipole field only causes transitions in the electronic shell for which $\Delta j = 1$, the quadripole for $\Delta j = 2$, and so on. Thus without any assumption as to the structure of the nucleus, we may state that the dipole field is the most general outgoing wave which causes S—P transitions, and so on. The dipole, quadripole, etc., waves are not special forms derived from a nuclear model.

On the other hand, there is no *a priori* reason why the wave radiated by the nucleus should not be the sum of terms of the type (5.2), (5.3), etc. Assuming a field of this type, one could obtain any value of the internal conversion coefficient. However, since we know (§ 4) that a given transition in atoms is associated with either a dipole, or quadripole, and since dipole and quadripole radiations are sufficient to explain the experimental results, it seems probable that nuclei also have this property.

§ 6. In view of the complicated nature of the calculations of § 2, and the result that for small $h\nu$ the quadripole internal conversion coefficient α is much greater than the dipole, it is of interest to obtain an analytical formula for the internal conversion coefficient by making a number of simplifying assumptions.

† Darwin, 'Trans. Camb. Phil. Soc.,' vol. 23, p. 137 (1924).

Such a formula may be in error by a factor of 3 or 4, but it will at least give us some idea of why the quadrupole α is bigger than the dipole, and of the results to expect for higher orders, octopole, etc.

The simplifying assumptions we make are the following: we use Schrödinger's equation; we assume the wave-length of the radiation long compared to the radius of the K ring (as in formula (3.1)). Also we represent the final state of the electron by a plane wave.

We consider the case where the angular momentum of the nucleus changes by n quanta. The Hertzian vector of the field of the nucleus is then

$$\mathbf{H} = 0, 0, B \frac{\partial^{n-1}}{\partial z^{n-1}} [(iqr)^{-1} \exp(iqr - 2\pi i \nu t) + \text{complex conjugate}]. \quad (6.1)$$

It is easily verified that the number of quanta radiated per unit time is

$$8\pi q^{2n-1} B^2 / \hbar (2n-1)(2n+1). \quad (6.2)$$

Assuming q small we obtain for A_0

$$A_0 = B \frac{\partial^n}{\partial z^n} \left(\frac{1}{iqr} \right) e^{-2\pi i \nu t} + \text{complex conjugate},$$

which reduces to (*cf.* Jeans, "Electricity and Magnetism," p. 215)

$$B n! r^{-n-1} (-1)^n P_n(\cos \theta) e^{-2\pi i \nu t} / iq.$$

The only final state to which transitions are possible has wave function, normalised as required in § 3,

$$v^{-\frac{1}{2}} P_n(\cos \theta) \left(\frac{2n+1}{4\pi} \right)^{\frac{1}{2}} \left(\frac{\pi k}{2r} \right)^{\frac{1}{2}} J_{n+\frac{1}{2}}(kr).$$

Thus the number of electrons ejected per unit time is

$$(a_0^3 \pi)^{-1} v^{-1} \frac{4\pi}{2n+1} \left(\frac{2\pi}{\hbar} \right)^2 \left| \varepsilon B n! \frac{1}{q} \int_0^\infty J_{n+\frac{1}{2}}(kr) r^{-n-1} e^{-r/a_0} r^2 dr \right|^2,$$

where a_0 is the radius of the K shell. To obtain an approximate expression we replace $J_{n+\frac{1}{2}}$ by the first term in the expansion of the Bessel function. We obtain for the number of electrons ejected per unit time

$$\frac{2\pi}{2n+1} \left[\frac{n!}{\Gamma(n+\frac{3}{2})} \right]^2 v^{-1} \left(\frac{2\pi \varepsilon}{\hbar} \right)^2 B^2 \frac{a_0^3 k^{2n+2}}{q^2 2^{2n+1}}.$$

Dividing by the number of light quanta we obtain for the internal conversion coefficient

$$\alpha = \frac{1}{2}\pi (2n + 1) \left[\frac{n!}{\Gamma(n + \frac{3}{2})} \right]^2 \frac{2\pi e^2}{h\nu} a_0 k \left(\frac{k}{2q} \right)^{2n+1}.$$

For long wave-lengths, $q \ll k$. Thus the larger the value of n the more rapidly will α increase for decreasing $h\nu$, a result which is borne out for $n = 1, 2$ by the exact calculations. Hence it is probable that for higher values of n , still larger values of α could be obtained.

In conclusion, we wish to express our thanks to Mr. H. R. Hulme, for allowing us to see his work in manuscript, and for his help and advice in the course of the calculations; to Dr. C. D. Ellis and Professor R. H. Fowler, for many discussions about the subject, and to Dr. Wishart, for the loan of a Brunsviga calculating machine, without which the computation would have been almost impossible.

Summary.

The internal conversion coefficient is calculated, assuming for the nucleus the field of a quadripole. Making use of Hulme's calculations with the dipole field, it is shown that the experimental results can be explained by the assumption that the nucleus can make transitions in which the angular momentum changes by one, or by two quanta. The bearing of this hypothesis on the theory of the nucleus is discussed.

The Rate of Burning of Colloidal Propellants.

By F. R. W. HUNT and G. H. HINDS, Military College of Science, Woolwich.

(Communicated by H. C. Plummer, F.R.S.—Received June 29, 1932.)

A.—Introductory.

1. Two papers have recently appeared in which a theoretical law is derived for the burning of colloidal propellants, and in each case the theory is supported by experimental results. Yamaga* proves that the rate of burning along the normal is proportional to the pressure of the surrounding gases, whereas Crow and Grimshaw† prove that it is proportional to their density. Further, Muraour‡ has shown by experiment that, with propellants containing a mineral jelly stabiliser, the rate of burning is proportional to the pressure, whereas with solventless propellants containing a centralite stabiliser, the rate of burning is a binomial function of the pressure, requiring the addition of a small constant term: this latter law Crow and Grimshaw§ have shown to be equivalent to their density law. With a view to resolving the differences between these conclusions, we have re-examined the experimental data given by Crow and Grimshaw.

2. But for the comparatively small effect of co-volume, the pressure of the gases would be proportional to the density; consequently a fairly delicate test is necessary to differentiate between the two rate-of-burning laws.

Various methods of examination were tried, including the following:—

- (a) A method depending on the maximum slope of the pressure-time curve. This was discarded because of the difficulty of measuring a maximum slope, and also because the influence of the co-volume on this slope was complicated and considerable.
- (b) Methods depending on the area under the pressure-time curve. These were discarded because small errors in the estimate of the time at which combustion is complete cause large errors in the deduced area.
- (c) A method whereby functions of the pressure were plotted against time; these functions were so chosen that, provided that combustion took

* 'Z. ges. Schiess-u. Sprengstoffw.,' vol. 25, p. 60 (1930).

† 'Phil. Trans.,' A, vol. 230, p. 387 (1932). This paper will be referred to subsequently as "C.P."

‡ 'C. R. Acad. Sci. Paris,' vol. 192, p. 227 (1931).

§ "C.P.," p. 397.

place in accordance with the law assumed, their graphs would be straight lines. This method was finally adopted because it enables a series of points on the experimental pressure-time curve to be used, and not merely the total area under the curve, nor the slope at any one particular point, and also because any departure from the rate-of-burning law, assumed to be true, is immediately revealed by a curvature in the line through the plotted points.

B.—The Functions of Pressure.

3. The equation of state of the uncooled gases in the explosion vessel is*

$$p = \frac{\lambda \Delta}{1 - \eta \Delta} \quad (1)$$

where

p is the gas pressure in kilograms per square centimetre at any moment,
 Δ is the density of the gases in grams per cubic centimetre at that moment,

η is the co-volume in cubic centimetres per gram, and

λ is the "force constant" of the propellant.

If C gm. of propellant, of density δ gr./c.c., be exploded in a vessel of capacity K c.c., then the density of the gases, at the moment when the fraction z of the charge has been burnt, is Δ where

$$\frac{1}{\Delta} = \frac{V}{z} + \frac{1}{\delta} \quad (2)$$

and

$$V = \frac{K}{C} - \frac{1}{\delta}.$$

Eliminating Δ between equations (1) and (2) gives

$$\frac{\lambda}{p} = \frac{1}{\Delta} - \eta = \frac{V}{z} + \frac{1}{\delta} - \eta. \quad (3)$$

For propellants in the form of long cylinders of the same diameter, we have the geometrical relation

$$z = 1 - f^2, \quad (4)$$

where f is the fraction of the diameter remaining at the moment when the fraction z of the charge has been burnt. This relation is based on Piobert's

* Crow and Grimshaw, 'Phil. Trans.,' A, vol. 230, p. 39 (1932). This paper will subsequently be referred to as "E.S."

law of burning by parallel layers, and neglects burning at the ends of the cylindrical pieces of propellant.

4. The two functions that we propose to use are

$$x = \log_e \frac{1+f}{1-f} + \frac{2f}{V\delta} \quad (5)$$

and

$$y = \log_e \frac{1+f}{1-f} + \frac{2f}{V} \left(\frac{1}{\delta} - \eta \right), \quad (6)$$

f being derived from the observed pressure by means of the relations

$$f^2 = 1 - z \quad (7)$$

and

$$\frac{V}{z} = \frac{\lambda}{p} - \frac{1}{\delta} + \eta. \quad (8)$$

Differentiating the first with respect to time, t , we have

$$\left. \begin{aligned} \frac{dx}{dt} &= 2 \left\{ \frac{1}{1-f^2} + \frac{1}{V\delta} \right\} \frac{df}{dt} \\ &= \frac{2}{V\Delta} \frac{df}{dt}. \end{aligned} \right\} \quad (9)$$

If therefore the rate of burning is proportional to the density of the gases, the graph of x plotted against t will be a straight line. In symbols, if

$$D \frac{df}{dt} = -u\Delta, \quad (10)$$

where D is the diameter of the cylinders of unburnt propellant in centimetres, and u is a constant depending on the nature of the propellant and on the charge temperature, then the slope of the line is given by

$$\frac{dx}{dt} = -\frac{2u}{V\Delta} \quad (11)$$

and the value of u can at once be derived.

Differentiating y with respect to time, we have

$$\begin{aligned} \frac{dy}{dt} &= 2 \left\{ \frac{1}{1-f^2} + \frac{1}{V} \left(\frac{1}{\delta} - \eta \right) \right\} \frac{df}{dt} \\ &= \frac{2}{V} \left\{ \frac{1}{\Delta} - \eta \right\} \frac{df}{dt} \end{aligned}$$

and, using (10),

$$\frac{dy}{dt} = -\frac{2u}{VD} \{1 - \eta\Delta\}. \quad (12)$$

Hence, initially, the slope of the graph of y , plotted against time, has practically the same value as that of the x graph; as burning progresses, however, Δ increases and the slope decreases in absolute value; the curve therefore bends upwards, and finally meets the x line on the time axis, since $x = y = 0$ when $f = 0$. If, therefore, the rate of burning is proportional to the density of the gases, the two functions, x and y , when plotted against time, will appear as indicated in fig. 1.

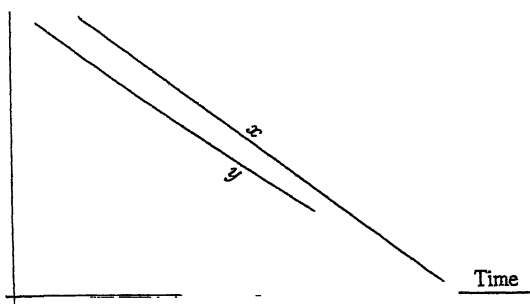


FIG. 1.—Density law.

If on the other hand the rate of burning is proportional to the pressure of the gases, such that

$$D \frac{df}{dt} = -qp. \quad (13)$$

where q is a constant depending on the nature of the propellant and on the charge temperature, then, from (12), (1), and (13),

$$\frac{dy}{dt} = \frac{2\lambda}{Vp} \frac{df}{dt} = -\frac{2\lambda q}{VD}, \quad (14)$$

and, from (9), (1) and (13),

$$\frac{dx}{dt} = -\frac{2\lambda q}{VD} \left\{ \frac{1 + \eta p}{\lambda} \right\}. \quad (15)$$

In these circumstances the graph of y plotted against time will be a straight line, while that of x will be a curve having approximately the same slope as that of y initially; as the pressure increases, the x curve will bend downwards, ultimately meeting the y line on the time axis. If, therefore, the rate of

burning is proportional to the pressure of the gases, the two functions, x and y , when plotted against time, will appear as indicated in fig. 2.

In the analysis which follows, each experimental pressure reading is corrected for cooling and other losses; from every such corrected reading, the corresponding values of the two functions, x and y , are calculated. We thus obtain quantities which depend on only one observation at a time, and are independent of previous and subsequent pressure observations. For each explosion, graphs of x and y against corresponding times are drawn.

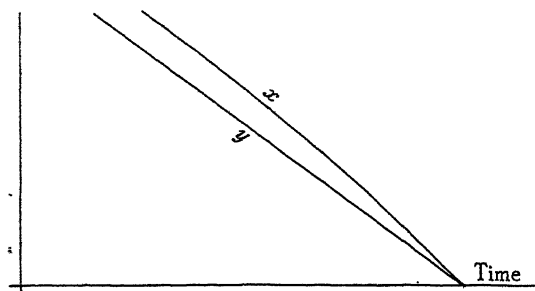


FIG. 2.—Pressure law.

C.—Analysis of Experimental Results.

5. *Correction for Energy Losses.*—The method developed by Crow and Grimshaw* of allowing for energy losses up to the instant of maximum pressure can be extended to allow for losses up to any time t during the explosion.

The work done in expanding the vessel is bp^2S , where b is a constant and S is the area of the internal surface of the vessel.

The rate of heat flow across unit area of the internal surface of the vessel at time t is $a(T - T_w)/2\sqrt{t}$, where T is the mean temperature of the gases at time t , T_w is the initial temperature of the internal surface of the vessel, and a depends on the properties of the steel of which the vessel is constructed.

Hence, if s is the mean specific heat of the gases, and T_0 is the uncooled temperature,

$$Cs\bar{s}(T_0 - T) = bp^2S + (T' - T_w)aS\sqrt{t},$$

where T' is the mean value of T with respect to \sqrt{t} . After the initial stage of combustion no serious error will be made in replacing T' by T in this equation.

The equation of state of the gases may be written in the form

$$p = \frac{T\lambda\Delta}{T_0(1 - \eta\Delta)}.$$

* "E.S.," pp. 50 *et seq.*

Combining this with equations (2), (4) and (10), and eliminating Δ , f and z , we obtain an equation of the form

$$\frac{pT_0}{\lambda T} = \Phi \left(\frac{ut}{D} \right).$$

The coefficient u depends essentially on the nature of the propellant, and is considered to be independent of the state of the gases; hence, at given time t , $pT_0/\lambda T$ is constant; consequently the densities in the actual and the uncooled states at time t are the same. It follows that p/p_0 is equal to T/T_0 , where p_0 is the uncooled pressure corresponding to p . In order then to correct observed pressures to uncooled pressures at the same instant, we have to apply Crow and Grimshaw's correction, modified to the extent of substituting Cz for C .

If the rate of burning is assumed to be proportional to the pressure instead of to the density, the same correction to observed pressure holds, provided that a small correction is also applied to observed times. Eliminating Δ , f and z from the above equation of state and equations (2), (4) and (13), we obtain an equation of the form

$$\frac{pT_0}{\lambda T} = \psi \left\{ \frac{\lambda q}{D} \cdot \frac{\bar{T}}{T_0} (t' - t) \right\},$$

where t' is the time at which combustion is complete, and \bar{T} is the mean value of temperature with respect to time during the time interval $t' - t$. Under uncooled conditions

$$\frac{p_0}{\lambda} = \psi \left\{ \frac{\lambda q}{D} \cdot (t'_0 - t_0) \right\},$$

where $t'_0 - t_0$ is the corrected value of $t' - t$. Hence

$$p_0 = \frac{T_0 p}{T}$$

when

$$t'_0 - t_0 = \frac{\bar{T}}{T} (t' - t).$$

6. *Application of Corrections.*—The experimental results given by Crow and Grimshaw* are time-intervals for increments of $1/50$ in the ratio p/p_m , where p is the manometer pressure, and p_m is the maximum manometer pressure, both uncorrected for losses due to the expansion of the explosion vessel, and

* "C.P.," p. 409.

to conduction through the walls. Corrections are applied by means of the formula given by them,* modified, as indicated above, to apply during com-

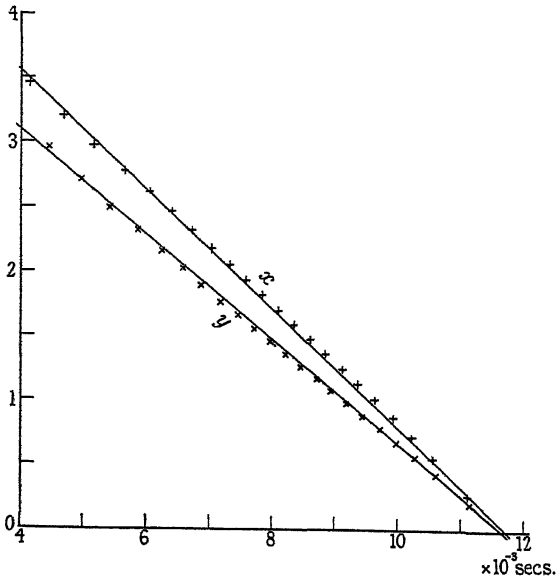


FIG. 3.—Series I. Propellant N.C. ; charge temperature 26.7°C .

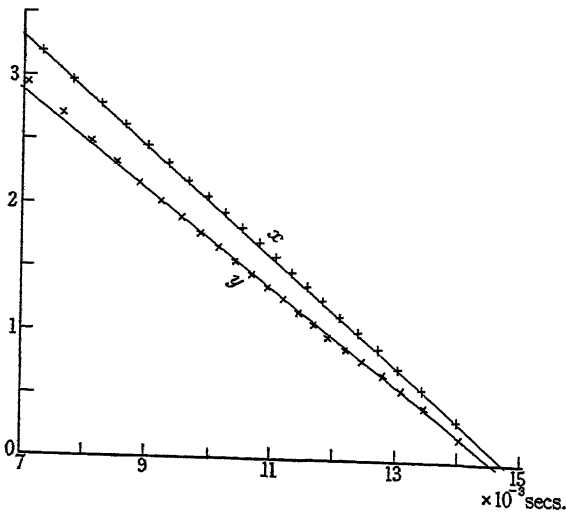


FIG. 4.—Series II. Propellant N.C. ; charge temperature -0.55°C .

bustion by the substitution of Cz for C. With this modification, and assuming that the correction is small so that the binomial approximation may be applied,

* "E.S.," equation (6A), p. 54.

the formula gives as the percentage additive correction to be applied to the experimental pressure

$$3.21 \times 10^{-4}p + 3000 \frac{\sqrt{t}}{Cz} + 1.984 \times 10^{-6} \frac{p^2}{Cz}$$

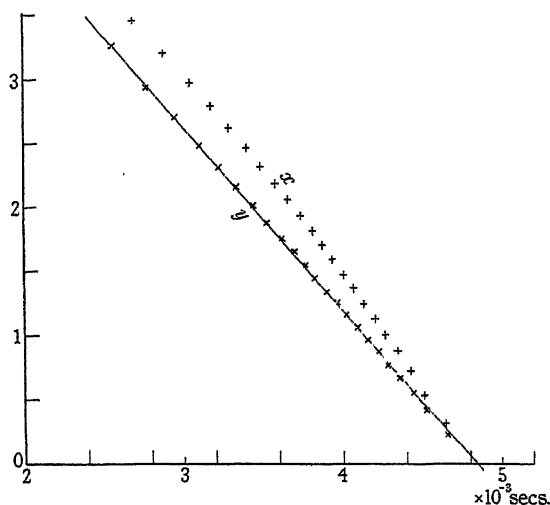


FIG. 5.—Series III. Propellant M.D.; charge temperature 26.0°C .

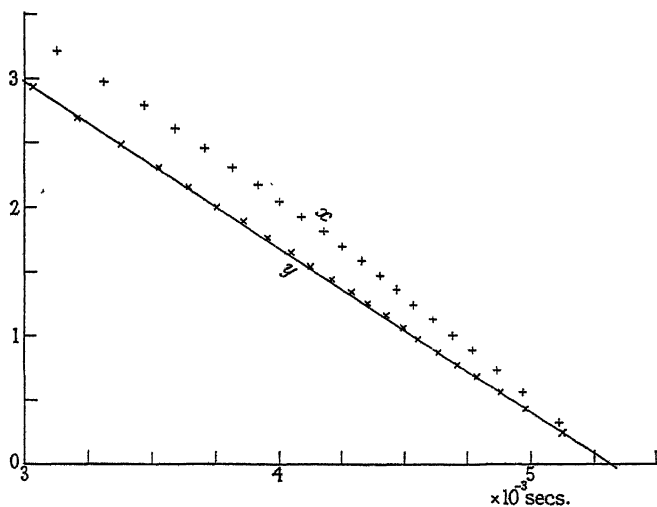


FIG. 6.—Series IV Propellant M.D.; charge temperature 1.67°C .

and for ease of manipulation this is applied in the form

$$0.321 \times 10^{-3}p + \frac{3000}{C} \frac{\sqrt{t \times 10^4}}{100z} + \frac{198.4}{C} \frac{(p \times 10^{-3})^2}{100z}$$

and z is taken, with sufficient accuracy for use in this formula, as equal to p/p_m . Zero time is difficult to arrive at, but is not required accurately for the purpose of applying the corrections, and is therefore obtained by extrapolation,

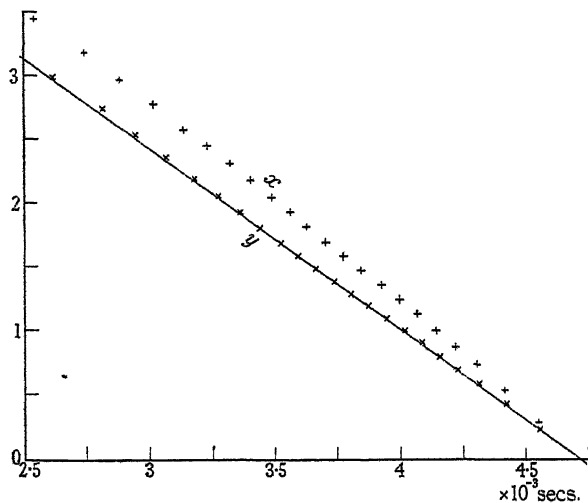


FIG. 7.—Series V. Propellant Mk. I; charge temperature 26.0°C .

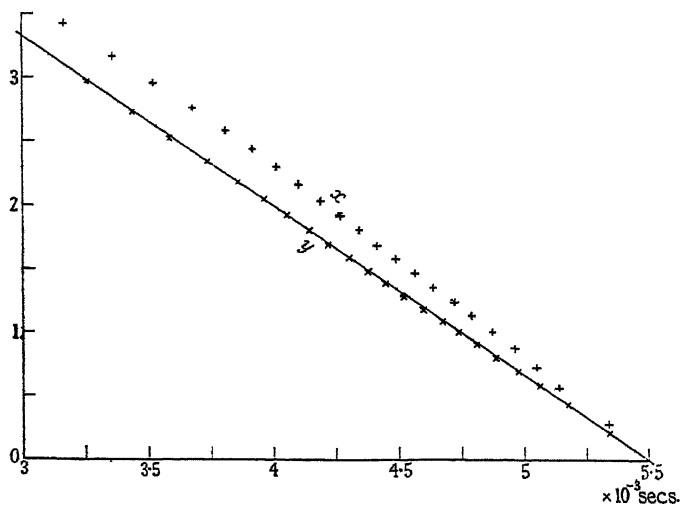


FIG. 8.—Series VI. Propellant Mk. I; charge temperature 0.55°C .

assuming that the pressure-time curve is linear from $p = 0$ to $p/p_m = 3/50$. From the pressures corrected in this way, values of x and y are calculated, and are plotted against corresponding times in the accompanying graphs, the time-co-ordinates of the y -points being corrected as indicated.

7. *Discussion of Results.*—In general, the smoothness of the curves bears testimony to the excellence of the experimental work. In all curves, the first two or three points do not conform, and this is explained by the fact that, at

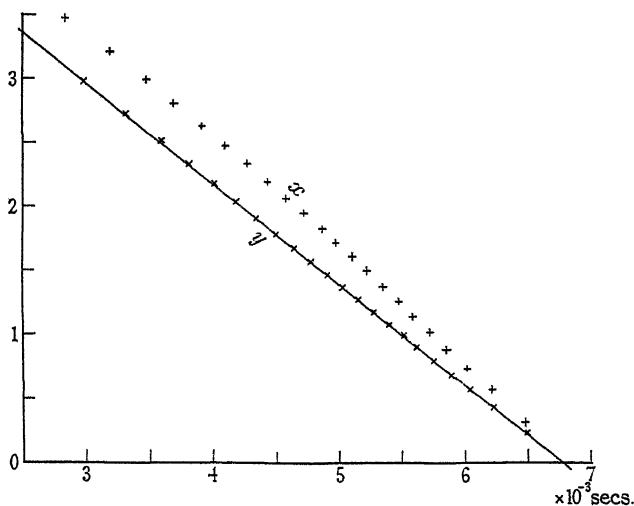


FIG. 9.—Series VII. Propellant H.C. ; charge temperature 26.7° C.

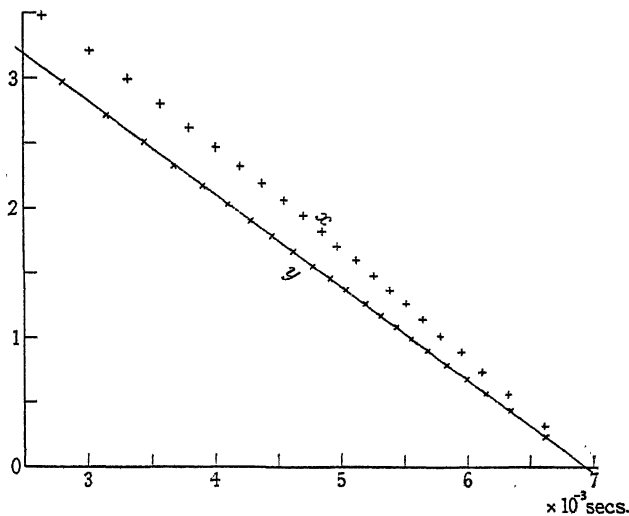


FIG. 10.—Series VIII. Propellant H.C. ; charge temperature 0.3° C.

low pressures, the sensitivity of the manometer is diminished,* and further, and probably more important, the correcting terms are of considerable magni-

* "E.S.," p. 58.

tude, so that the approximations used in applying the correction formula no longer hold good ; for this reason the first two or three points are neglected.

From the experiments with the N.C.* propellant, series I (charge temperature $26\cdot6^{\circ}\text{C.}$), the x curve tends to bend downwards towards the end of combustion, while the y curve tends to bend upwards : no conclusive deduction can be drawn from this series. In series II, the x curve is definitely more nearly straight than the y curve, indicating a rate of burning proportional to density.

In all other series, the y curve is straight, and the x curve shows the characteristic downward curvature to be expected if the rate of burning is proportional to pressure.

It thus transpires that, whereas Crow and Grimshaw deduce a density law in all cases, we find evidence of it only with the N.C. propellant ; in the M.D., Mk. I, and H.C. propellants, the evidence is that the rate of burning is proportional to pressure. This conclusion is to a certain extent similar to that of Muraour (*loc. cit.*) who found the pressure law to hold in the case of propellants containing a mineral jelly stabiliser.

8. *Rate-of-Burning Constant.*—In the case of propellants other than N.C., the value of the constant, q (equation (13)), can be immediately derived from the y curves : in the case of N.C., a best straight line can be drawn, and the corresponding values of q obtained. These values of q are collected in the following table :—

Series.	Propellant.	Charge temperature.	Loading density.	q .
		$^{\circ}\text{C.}$		
I	N.C.	26·7	0·2479	0·0141
II	N.C.	—0·55	0·2479	0·0133
III	M.D.	26·0	0·2479	0·0202
IV	M.D.	1·67	0·2479	0·0182
V	Mk. I	26·0	0·2211	0·0298
VI	Mk. I	0·55	0·2211	0·0281
VII	H.C.	26·7	0·2479	0·0272
VIII	H.C.	0·30	0·2479	0·0248

9. In conclusion, we wish to express our thanks to Dr. A. D. Crow for helpful

* The symbols used here for the various propellants are those adopted by Crow and Grimshaw ('C.P.', p. 404), *i.e.*, N.C. is a straight nitrocellulose propellant stabilised with diphenylamine ; M.D. is Service cordite (a cool nitrocellulose-nitroglycerine mixture stabilised with mineral jelly) ; Mk. I is Service cordite Mark I (a hot nitrocellulose-nitroglycerine mixture stabilised with mineral jelly) ; and H.C. is a hot nitrocellulose-nitroglycerine mixture stabilised with centralite.

criticism, and to the Ordnance Committee and to the Commandant of the Military College of Science for permitting publication.

Summary.

A method of analysing pressure-time curves obtained from closed-vessel experiments is developed, whereby two functions of the pressure are plotted against time. One yields a straight line if the rate of burning is proportional to the density, while the graph of the other is straight if the rate is proportional to the pressure. Only one of the four propellants considered by Crow and Grimshaw ('Phil. Trans.,' A, vol. 691 (1932)) gives evidence of a departure from the pressure law.

The Relationship between Viscosity, Elasticity and Plastic Strength of Soft Materials as Illustrated by some Mechanical Properties of Flour Doughs, I.

By ROBERT KENWORTHY SCHOFIELD and GEORGE WILLIAM SCOTT BLAIR.

(Communicated by Sir John Russell, F.R.S.—Received July 8, 1932.)

[PLATE 15.]

1. Flour dough belongs to a group of materials in which a high degree of plasticity is combined with considerable elasticity. Owing to their great industrial importance, a number of technological investigations have been carried out on these materials (of which unvulcanised rubber is another important example); but the problem of bringing the description of their behaviour within the scope of a general theory of viscosity and elasticity has hardly been tackled.

The time during which a stress is applied is as important as the magnitude of the stress in determining how much of the deformation is elastic (recoverable) and how much plastic (non-recoverable). This fact suggests that a formulation based on Maxwell's "time of relaxation" should be of value in this connection. The formulation as Maxwell gave it applies to a true fluid, for which case the relaxation is exponential and for which the rate of dissipation of internal

stress is proportional to the stress, the constant of proportionality being the reciprocal of the relaxation time. Thus

$$-\frac{ds}{dt} = \frac{1}{t_r} S,$$

or

$$\frac{1}{t_r} = -\frac{1}{S} \frac{dS}{dt} = -\frac{d(\log_e S)}{dt},$$

where S is the shearing stress, and t_r the relaxation time.

An obvious method of extending this conception to plastic materials is to recognise the relaxation time as a quantity which is not a constant but which varies according to the stress, and to consider its reciprocal as determined by $-\frac{d(\log_e S)}{dt}$. Thus the reaction of the material to an external stress will be more purely elastic in proportion as the time of relaxation exceeds the time of application.

In handling a flour dough, strains up to 30 per cent. given momentarily recover almost completely, so that, for the corresponding stresses, the time of relaxation must be large enough to be measured without the aid of elaborate apparatus. The reason for this is brought out by a further consideration of Maxwell's theory. In steady flow the velocity gradient (or rate of change of shearing strain), G , is related to the rate of dissipation of internal stress by the rigidity modulus, n , of the material. Thus

$$-\frac{dS}{dt} = nG.$$

So that the viscosity, η , which is defined as the ratio of S to G is given by

$$\eta = nt_s. \quad (1)$$

Hence

$$t_r = \eta/n.$$

In the case of plastic materials, η , if defined as the ratio of S to G , is not a constant,* but varies with S ; n , on the other hand, appears to be a constant of

* There is no universally recognised definition of viscosity in systems other than true fluids. In certain cases, although S/G varies, dS/dG is constant over a considerable range of stress. In such cases the behaviour of the system is conveniently described in terms of dS/dG which has been termed the "pseudo-viscosity"† or "stiffness," its reciprocal being Bingham's "mobility."‡ It seems best to reserve the term viscosity for the simple ratio S/G , and recognise that, in general, this varies with S .

† Scott Blair and Crowther, 'J. Phys. Chem.,' vol. 33, p. 321 (1929).

‡ Bingham, "Fluidity and Plasticity" (1922).

the material. Under low stresses the value of η for flour dough is considerable, while the rigidity modulus is extremely small by comparison with that of ordinary solids. It is the combination of these extreme properties which gives this material a special interest, and makes its study valuable for the purpose of advancing our general understanding of these systems.

2. Experiments which have enabled a direct determination to be made of t_r are dealt with in Section 3. It is convenient, however, before describing these to outline the results of some preliminary experiments which bring out the importance of the time of application in determining the type of reaction to an external stress. The first experiments were carried out with the pachimeter, an instrument for measuring plastic strength under rolling which has already been described in detail in its application to the study of soils.* A test piece in the form of a small cylinder of known radius and length is made to roll between two horizontal plates by reciprocating the lower one, and the instrument is designed to measure the stress which must be exerted by the upper plate in order to produce in the rolling cylinder a permanent elongation.

It was thought that the plastic strength of a dough thus determined might have an important bearing on the baking quality of the flour. It was found that the instrument is capable of distinguishing between flours, but the results did not run entirely parallel with the bakers' opinion. It is now clear that at least one reason for this is that, while the stresses set up by the gas generated in a yeasted dough operate over a time to be measured in hours, the reciprocating action in the pachimeter impresses a stress on any given diameter of the rolling dough cylinder for only a fraction of a second. The fact that a dough cylinder does not show a measurable permanent elongation until a stress greater than a certain critical stress has been applied between the plates must be interpreted as showing that below this critical stress the time of relaxation is too long for any measurable plastic deformation to occur in the time allowed by the instrument. The remarkable reproducibility of the figures given by this instrument is evidently due, in part, to the fixing of the time of application of the stress by the speed of reciprocation; but it must also, in part, be attributed to the rolling action, through which every diameter comes under stress in turn thereby averaging out the effects of irregularities in the specimen. It is evident, however, that for such a material the terms *plastic strength*, and *yield value*, if used to describe the figures obtained, must be construed as

* Schofield and Scott Blair, 'J. Agric. Sci.', vol. 132, p. 135 (1932); 'J. Soc. Chem. Ind.', vol. 51, p. 205 (1932); 'Trans. Ceram. Soc.', vol. 31, p. 79 (1932).

relative rather than absolute terms. A second series of measurements was made on a newly designed instrument which we have called a rack owing to its superficial resemblance to an ancient instrument of torture, fig. 8, Plate 15. Long cylindrical pieces of dough (made by forcing the material through a short piece of metal tubing attached to the body of a grease gun) were stretched out and held stretched for a measured time, at the end of which they were cut loose and their elastic contraction measured.*

Fig. 1 shows the result of a series of experiments in which cylinders of a dough† of a good bakers' mixture were stretched to different extents and cut loose after 1 minute. It will be seen that the elastic recovery, expressed as

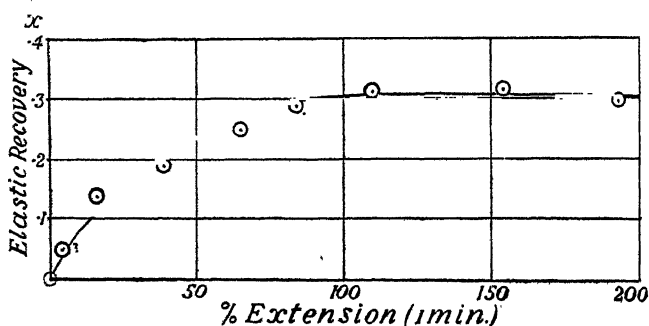


FIG. 1.

the ratio of the contraction to the final length, increased as the initial extension was increased to 100 per cent., but that no further increase was obtained up to 200 per cent. A curve resembling the earlier portion would be obtained on plotting the same quantities for a copper wire as may be seen from the data recently published by Taylor and Quinney.‡ It is not suggested that the correspondence is complete, but a due recognition of the widely different relaxation times involved in the two cases does much to bridge the gap. In a further series of measurements cylinders made from portions broken from a large dough at intervals during a period of several hours were stretched to

* The satisfactory agreement obtained between duplicate dough cylinders made up by this method demonstrates that any irregular strains set up in the dough in the grease gun are not great enough to upset the behaviour of the cylinder when under stress.

† In comparing doughs made from different flours it was not convenient to use in every case the same proportion of flour to water. It was found possible to reproduce dough samples most satisfactorily by making them up to that moisture content at which they will just not stick to a glass plate when pressed firmly on to its surface. All doughs were made up with 23 per cent. salt solution.

‡ 'Phil. Trans.,' A, vol. 230, p. 323 (1931).

just over 100 per cent. and their elastic recovery after 1 minute was measured. No appreciable differences were found. These results are in striking contrast to those obtained from portions of the same dough and investigated in the pachimeter. As will be seen from fig. 2 the pachimeter-values (W) show

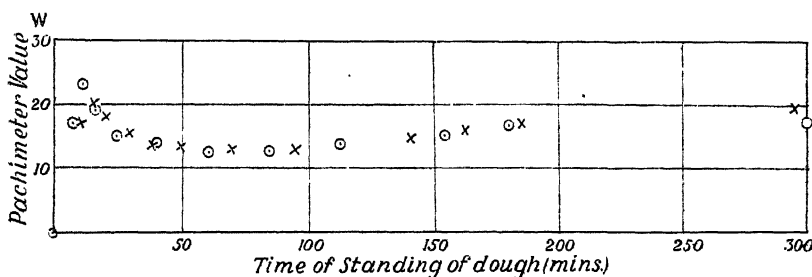


FIG. 2.—○ × duplicate doughs.

marked fluctuations with time. In order to ascertain whether the results from the two instruments differ because of a relative shortness of time under stress in the pachimeter, a simple experiment was made. The apparatus constructed consisted essentially of a ballistic pendulum which, being released from an angle of 25° , was allowed, when at the lowest point of its swing, to strike and rebound from the circular face of a cylinder of dough. The cylinder had a radius of 3 mm. and length 15 mm. and was held in an L-shaped support so that the end away from the pendulum remained rigid and the cylinder was given support from below for about three-quarters of its length. The pendulum bob consisted of a clay marble weighing 2.3 gm. supported by two threads

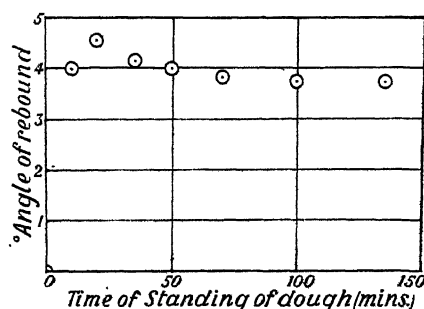


FIG. 3.

45 cm. long separated by 15 cm. at their upper end. The results, shown in fig. 3 (each point represents a mean of three readings) confirm the expectation that the larger values of W , since they indicate a smaller dissipation of rapidly applied stresses, correspond to bigger angles of rebound.

Although this is not the place to discuss how far the variations in W reflect the colloidal and other changes which take place between the time when a baker first mixes his dough and that when it is "ripe" for baking, it is interesting to note that there is a correspondence between them and the variation of consistency of flour pastes as recorded by Jago for doughs* and by Denham, Scott Blair and Watts and others† for flour pastes. Since these consistency measurements were made by observation of fairly rapid flow through capillary tubes, it is satisfactory to note that agreement is shown with the pachimeter in which shearing is rapid rather than with the rack where the relaxation of stress is slow.

3. The rack may also be used to determine the time of relaxation for a certain range of stresses. For if, as seems reasonable, we may assume a constant proportionality to exist between the elastic recovery, x , and the stress still undissipated at the time of cutting loose, the reciprocal of the time of relaxation is as well given by $-d(\log_e x)/dt$ as by $-d(\log_e s)/dt$.‡ The former quantity can be evaluated by obtaining from a series of experiments, in which the initial deformation is the same, the variation of x with the time during which the cylinder is held deformed. The result of such a series is shown in fig. 4, the circles giving the values of x and the crosses of $\log_e x$. Fig. 5 gives

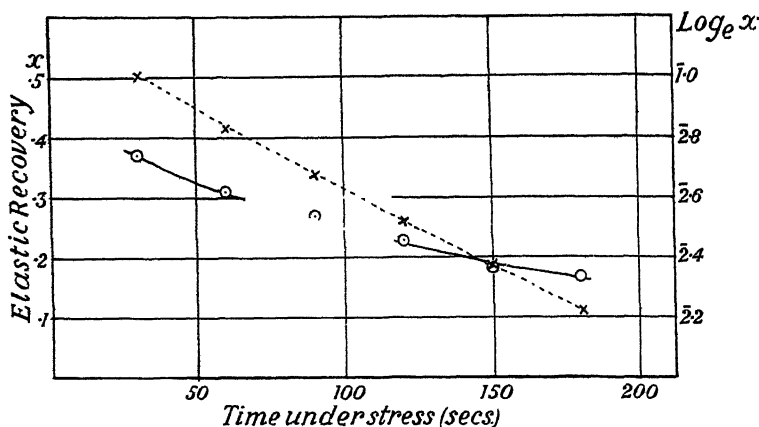


FIG. 4. $\circ = x$, $\times = \log_e x$.

* Jago, "Chemistry of Wheat, Flour, and Bread, and Technology of Bread Making" (1886 and later editions).

† Denham, Scott Blair and Watts, 'Cereal Chem.', vol. 4, p. 206 (1927); Sharp and Gortner, 'Tech. Bull. Minn. Agr. Exp. Sta., No. 19,' (1923); St. John and Bailey, 'Cereal Chem.', vol. 4, p. 140 (1929).

‡ Where s is the tensile stress per unit area (equals $3S$, see below), the stress plotted in figs. 5, 6 and 7 is the tensile stress, s .

the variation of the time of relaxation with the stress. In order to deduce the undissipated stress plotted as abscissa it was necessary to obtain the appropriate value for Young's modulus. For this purpose observations were made on the compression caused by placing a small weight on top of a squat cylinder of dough 1 cm. high and 0.55 cm. mean diameter. A fourfold magnification was obtained in measuring the deformation by an arrangement resembling the backsight of a rifle. This method was used because it was simple, and enabled the readings to be taken in quick succession, so that the cylinder was not loaded long enough for any appreciable plastic flow to take place, fig. 9,

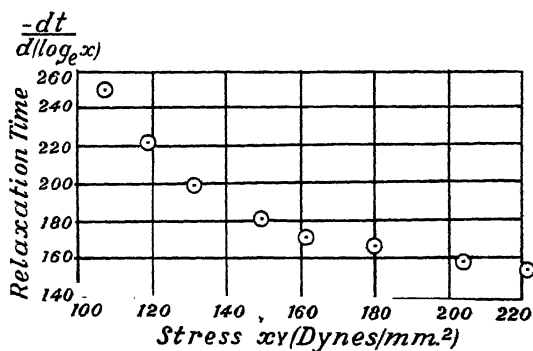


FIG. 5.

Plate 15. A mean value of about 4×10^4 dynes per square centimetre was obtained, and no variations could be detected during the ageing of a dough. This figure is admittedly only approximate, as, apart from the smallness of the deformation to be measured, it was difficult to form the dough into a true cylinder and obtain a satisfactory value for its cross-section. On the other hand, it is certainly not seriously in error; but since higher values were obtained by another method, to be described later, a somewhat higher value, namely, 6×10^4 has been adopted.

In order to obtain a check on the general correctness of fig. 5 and also to extend the stress range of the determination of the time of relaxation, a modification of the rack was constructed in which the weight of the dough cylinder was supported on a pool of mercury contained in a long shallow wooden trough. One end of the dough was fixed by pressing it round a screw let into one end of the trough while the other was attached to a thin strand of rubber by the extension of which the stress on the dough could be observed. The other end of the rubber strand was secured to a thread which could be wound up on a small winch. Direct observations on the decay of the stress were made by

first rapidly extending the dough by winding the thread on to the winch until after an extension of about 150 per cent. an ink mark on the dough came into the field of view of a low-power microscope. The thread was then gradually released at rate carefully regulated so as to keep the ink mark on the cross-wire of the microscope. From a separate calibration the stress corresponding to each notch was known, and this, divided by the mean cross-section, gave the ordinate of fig. 6, the abscissa being the recorded times. In fig. 7 is given

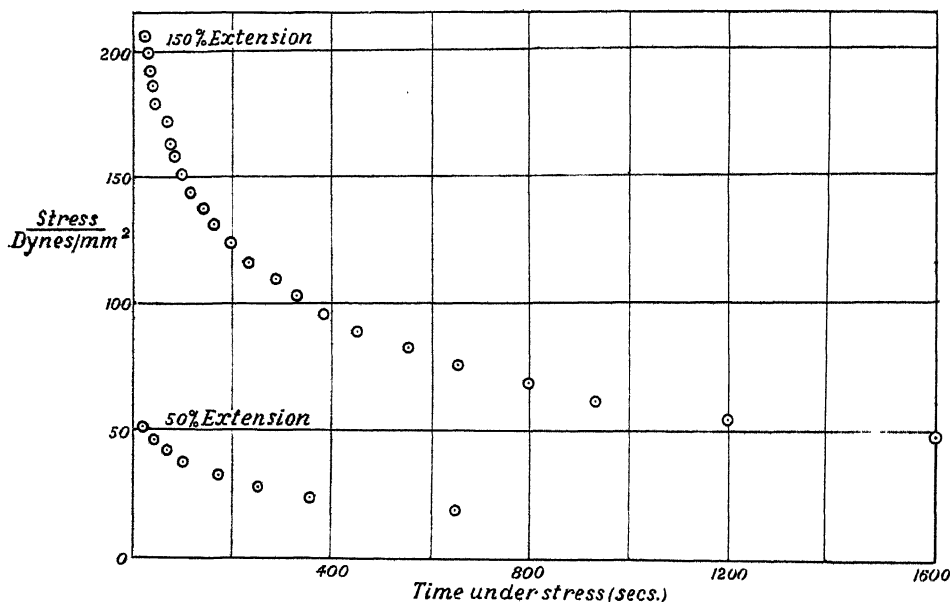


FIG. 6.

the curve for the relaxation time derived from it. It shares with fig. 5 the uncertainty attached to the evaluation of a cross-section, but is in other respects more reliable and extends over a much wider range.

In figs. 6 and 7 are also shown curves relating to the decay of stress for an initial deformation of only 50 per cent. The lack of agreement between the two curves in fig. 7 shows that the time of relaxation depends on the deformation the specimen has suffered as well as on the stress. The dependence of time of relaxation (and hence of viscosity) on the total deformation was suspected from the results of some rougher comparisons of a similar nature made with the original rack. It has been confirmed in a new series of experiments in which observations have been made of the plastic extension of dough cylinders under their own weight, an account of which will be published shortly.

In these experiments it has been found that although the cylinders thin considerably in extending there is no corresponding increase in the rate of extension. This phenomenon may be compared on the one hand with the hardening of metals under working and on the other with an analogous effect observed by Trouton in the case of pitch.* These results point out a weakness† in the usual methods of measuring viscosity by observations on steady flow. Unless special care is taken the only value obtained for the viscosity will be the limiting value for large total deformations.

In a complete study of the plastic behaviour of such a material the viscosity-stress relationship should be found for a series of deformations. The measure-

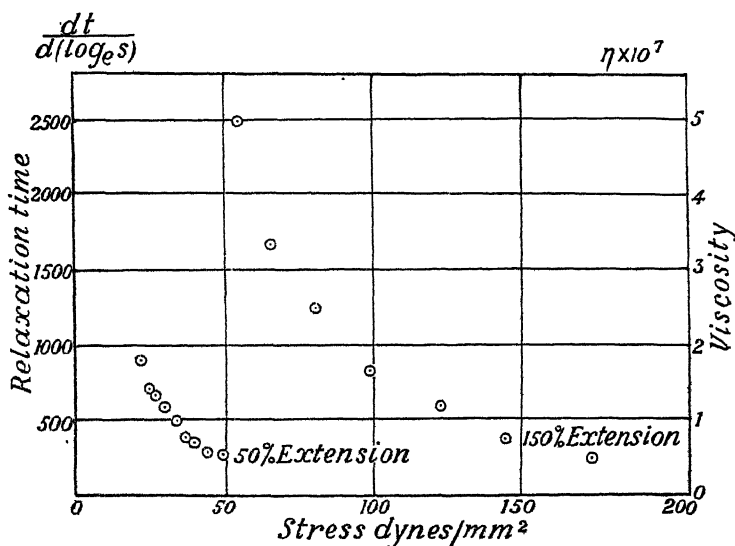


FIG. 7.

ment of flow at constant deformation may seem at first sight to be impossible, but this is actually what was done in the experiments just described. While the winch remained on a given notch the dough was flowing slowly under a substantially constant stress. Before it had gone far, however, the stress was reduced and an elastic recovery occurred equal and opposite to the deformation

* 'Proc. Roy. Soc.,' A, vol. 77, p. 426 (1906).

† The determination of viscosity by the usual methods is also liable to disturbance from slip or anomalous flow at the solid surfaces of the viscometer (Schofield, R. K., and Scott. Blair, G. W., 'J. Phys. Chem.,' vol. 34, p. 248, 1505 (1930); 'J. Phys. Chem.,' vol. 35, p. 1212 (1931); 'Phil. Mag.,' vol. 72, p. 890 (1931)). In the method here described no such complication can arise.

caused by the flow. This deformation can therefore be obtained by dividing the stress change by Young's modulus.

According to well-known principles a material such as this for which Poisson's ratio may be taken as $\frac{1}{2}$, when elongated may be considered as subject to two shearing strains each equal in magnitude to the fractional elongation. The corresponding shearing stresses are each one-third of the normal tensile stresses applied in causing the elongation, so that by plotting the rate of shear $-\frac{1}{Y} \frac{ds}{dt}$ against the shearing stress, $\frac{1}{3}s$, a flow-curve is obtained. The viscosity corresponding to any given stress s is therefore given by

$$\eta = -\frac{1}{3}s \left/ \frac{1}{Y} \frac{ds}{dt} \right.,$$

but

$$t_r = -s \left/ \frac{ds}{dt} \right..$$

Therefore

$$\eta = \frac{1}{3}Y t_r.$$

This result might have been deduced directly from equation (1) (p. 708), since when Poisson's ratio is $\frac{1}{2}$, $Y = 3n$; but its independent deduction serves a useful purpose if it helps to make clear how viscosity can be measured, in the case of a markedly elastic material, at constant deformation. The change which goes on in a dough held at constant deformation involves an increase of the plastic part of the deformation at the expense of the elastic part; and the rate of increase of the plastic deformation, or flow, is therefore known, if the rate of decrease of the elastic deformation has been obtained, since the sum of the deformations remains constant.

As a check on the value of the Young's modulus, and to test the correctness of using it to relate the undissipated stress to the elastic recovery, an experiment, which was started in the same way as those already described, was carried out on the modified rack. After keeping the ink mark on the cross-wire of the microscope for about a minute, the stress was completely released and the elastic recovery observed. The magnitude of the stress just before it was released having also been observed, it was only necessary to obtain the mean cross-section in order to evaluate Y . The elongation of the cylinders when floated on mercury is not quite so even as when they are supported on rubber bands, and so the cross-section could not be determined very exactly. A value of 6×10^4 was obtained. It has been preferred to the value 4×10^4 obtained with the earlier method as any unevenness in the faces of the loaded

cylinder would cause an apparent increase in the deformation thereby giving too low a value for Y . It is to be understood, however, that this value is provisional in view of the difficulties of the experiment. The precise numerical value does not affect the general form of fig. 5.

Discussion.

4. In the foregoing sections it has been shown how the conception of the time of relaxation, which was used by Maxwell as a method of describing the viscous behaviour of a liquid, can be extended to cover the behaviour of a complex system such as a flour dough. An attempt will now be made to specify the type of internal structure indicated by the results of the investigation.

Firstly, it is clear that the dough contains elastic elements which form a connected structure and that the determination of Young's modulus has reference to these elements. Secondly, it is evident that the elements, though connected, are not joined securely, but slide past one another whenever a sufficient stress is operative. The viscosity which has been determined is mainly governed by the behaviour of a plastic film by which the elastic elements are connected. It is quite possible that the elements are capable of complete elastic recovery, but there is at present no criterion for testing this. The time of relaxation is a characteristic of the connected structure as a whole, and its value is as much determined by the elasticity of the elements as by the viscosity associated with their plastic junctions.

The considerable time (several minutes) which often elapses between the release of the stress and the cessation of contraction indicates the existence of another viscosity* which must be distinguished from that already considered. The fact that the dough exhibits elastic recovery at all shows that this second viscosity is of a relatively low order, for if the viscous resistance to change of

* This second viscosity appears to be of the same nature as that discussed by Shorter in his explanation of the slow extension of hair and wool fibres.† "The fibre contains elastic elements with very different degrees of damping, so that on the first application of an external force the more lightly damped elements extend, and, as time goes on, the extension of the more highly damped elements begins to show itself." And again, "We get (where a fibre is held stretched to a definite length) an apparent elastic relaxation, which, however, is very different from the effect contemplated in Maxwell's theory of viscosity. It is not the disappearance of a state of strain owing to molecular readjustments, it is merely the transference of a state of strain from lightly damped to highly damped elements."

† Shorter, S. A., 'J. Text. Inst.,' vol. T. 15, p. 207 (1924); 'Trans. Faraday Soc.,' vol. 20, p. 228 (1924); 'J. Soc. Dy. Col. Bradford,' vol. 41, p. 212 (1925).

shape were greater than that involved in the dissipation of internal stress, the stress would be dissipated before the dough had made any appreciable recovery. The second viscosity may be associated with the medium in which the elastic elements are embedded, but we cannot rule out the possibility of its being somehow connected with the elastic elements and their system of connection.

In relating these deductions to the known structure of the dough one may safely identify the elastic elements with the protein part of the flour. That the elements form a connected structure is confirmed by the fact that the starch and the other constituents of the flour can be washed out of the dough without breaking up the gluten. That the slowness of the elastic recovery is partly due to the presence of the starchy aqueous medium in the dough is indicated by the fact that the elastic recovery is more rapid in washed gluten than in the dough itself. The recovery of washed gluten is, however, not instantaneous, so that part of the second viscosity must be attributed to the gluten.

Our thanks are due to Professor G. I. Taylor, F.R.S., for his advice on the choice of data for inclusion in this paper, and to Dr. B. A. Keen for helpful criticism; also to Dr. E. A. Fisher, Director, and to Dr. P. Halton of the Research Association of British Flour Millers who have kindly provided the flours used in this work, and have given us much useful technical information.

Summary.

(1) An extended significance is given to Maxwell's "time of relaxation" and this has been used in quantitatively describing the viscous and elastic behaviour of flour dough.

(2) The length of the time of application of a stress in relation to the corresponding time of relaxation determines what proportion of the deformation is elastic (recoverable) and what proportion plastic (non-recoverable).

(3) This fact is illustrated by a comparison of the behaviour of dough in the "pachimeter" and on the "rack," the behaviour in the "pachimeter" (rapid stressing) being paralleled by that exhibited in a ballistic experiment.

(4) The decay of internal stress in pieces of dough which had been stretched out and held stretched has been followed, and the times of relaxation, and the corresponding viscosities have been evaluated for a series of stresses.

(5) Dough shows a phenomenon similar to the hardening of metals under working as a result of which the time of relaxation and the viscosity for a given stress depend on the total deformation.

The internal structure of the dough thus revealed is briefly considered.

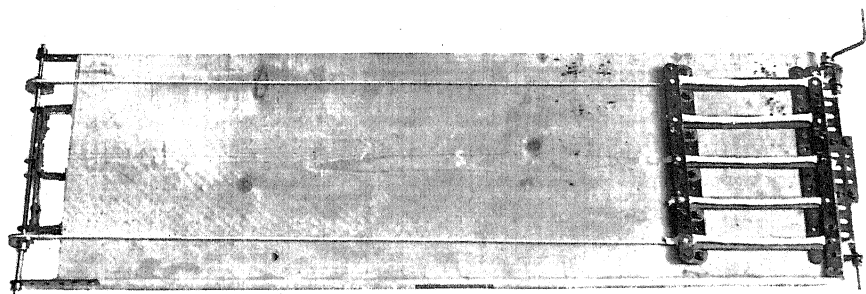


FIG. 8.

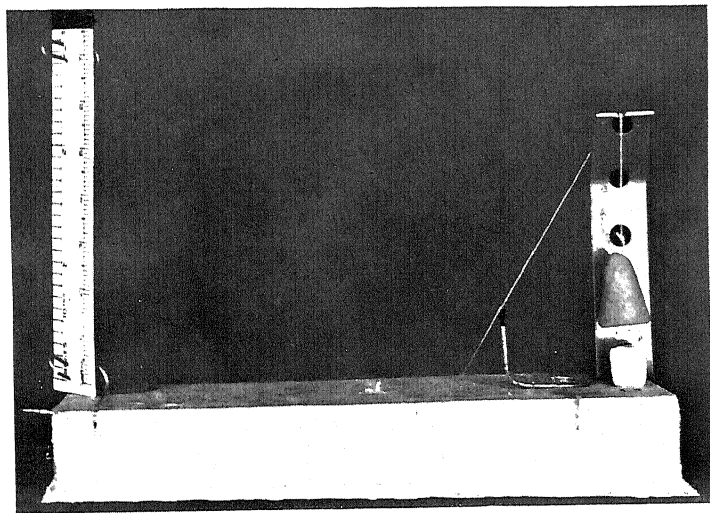


FIG. 9.

INDEX to VOL. CXXXVIII. (A.)

Adam (N. K.) and Harding (J. B.) The Structure of Surface Films, XVI, 411.

Adsorption (Hardy and Nottage), 259.

Andrade (E. N. da C.) and Chalmers (B.) The Resistivity of Polycrystalline Wires in Relation to Plastic Deformation, and the Mechanism of Plastic Flow, 348.

Badami (J. S.) *See* Rao and Badami.

Bailey (C. R.) *See* Cassie and Bailey.

Bangham (D. H.), Fakhoury (N.) and Mohamed (A. F.) The Swelling of Charcoal.—II, 162.

Blair (G. W. Scott) *See* Schofield and Blair.

Brata (L.) *See* Powell and Brata.

Burton (A. C.) *See* McLennan and others.

Butler (J. A. V.) *See* Hamilton and Butler.

Carbon monoxide, explosive oxidation at lower pressures (Hadman and others), 297.

Cassie (A. B. D.) and Bailey (C. R.) Investigations in the Infra-red Region of the Spectrum.—VII, 531.

Chain reaction between hydrogen and oxygen, lower pressure limit (Hinshelwood and Hughes), 311.

Chalmers (B.) *See* Andrade and Chalmers.

Charcoal, swelling (Bangham and others), 162.

Colloidal propellants, rate of burning (Hunt and Hinds), 696.

Davidson (P. M.) Rotational Uncoupling, with application to the Singlet Hydrogen Bands, 580.

Dipole moments of sulphur oxychlorides (Smith), 154.

Dutta (A. K.) The Absorption Spectrum of Nitrous-Oxide and the Heat of Dissociation of Nitrogen, 84.

Dye (the late W. D.) The Modes of Vibration of Quartz Piezo-Electric Plates as Revealed by an Interferometer, 1.

Eddington (Sir Arthur) Theory of Electric Charge, 17.

Elasticity of Soft Materials (Schofield and Blair), 707.

Electric charge, theory (Eddington), 17.

Electrode potential characteristic of metal (McAulay and Spooner), 494.

Electrode potential, mechanism (McAulay and Spooner), 494.

Electron scattering in gases, I (Mohr and Nicoll), 229.

Electrons, scattering in gases, II, large angle (Mohr and Nicoll), 469.

Electrolytes, behaviour in mixed solvents (Hamilton and Butler), 450.

Ellis (C. D.) The γ -Rays of Thorium B and of the Thorium C Bodies, 318.

Endemicity, problem of (Kermack and McKendrick), 55.

Epidemics, mathematical theory (Kermack and McKendrick), 55.

Errors and least squares, theory of (Jeffreys), 48.
Explosives, mechanism of detonation in solid (Taylor and Weale), 92.

Fakhoury (N.) *See* Bangham and others.

Films, structure of surface (Adam and Harding), 411.

Films, unimolecular, surface potentials (Schulman and Hughes), 430.

Gamma-rays, theory of internal conversion (Taylor and Mott), 665.

Gamma-rays of thorium B and C (Ellis), 318.

Hadman (G.), Thompson (H. W.) and Hinshelwood (C. N.) The Explosive Oxidation of Carbon Monoxide at Lower-Pressures, 297.

Halliday (E. C.) The Polarity of Thunderclouds, 205.

Hamilton (R. T.) and Butler (J. A. V.) The Behaviour of Electrolytes in Mixed Solvents, IV, 450.

Harding (J. B.) *See* Adam and Harding.

Hardy (Sir William) and Nottage (M.) Adsorption. A Study of Availability and Accessibility, 259.

Havelock (T. H.) The Theory of Wave Resistance, 339.

Hinds (G. H.) *See* Hunt and Hinds.

Hinshelwood (C. N.) *See* Hadman and others.

Hinshelwood (C. N.) and Moelwyn-Hughes (E. A.) The Lower Pressure Limit in the Chain Reaction between Hydrogen and Oxygen, 311.

Hughes (A. H.) *See* Schulman and Hughes.

Hughes-Moelwyn (E. A.) *See* Hinshelwood and Moelwyn-Hughes.

Hulme (H. R.) The Internal Conversion Coefficient for Radium C, 643.

Hunt (F. R. W.) and Hinds (G. H.) The Rate of Burning of Colloidal Propellants, 696.

Hydrogen bands, rotational uncoupling (Davidson), 580.

Hydrogen, formulæ for stark effect (MacDonald), 183.

Hydrogen molecule, the $2p^3$ II bands (Sandeman), 395.

Hydrogen, resonance spectrum (Rao and Badami), 540.

Ions in gases, mobility of alkali (Powell and Brata), 117.

Jackson (C. V.) Interferometric Measurements in the Spectrum of Krypton, 147.

Jeffreys (H.) On Plasticity and Creep in Solids, 283.

Jeffreys (H.) On the Theory of Errors and Least Squares, 48.

Kermack (W. O.) and McKendrick (A. G.) Contributions to the Mathematical Theory of Epidemics, II, 55.

Krypton, interferometric measurements of spectrum (Jackson), 147.

Leavey (E. W. L.) *See* Shaw and Leavey.

Ludlam (E. B.) *See* Ritchie and Ludlam.

McAulay (A. L.) and Spooner (E. C. R.) A Unique Electrode Potential Characteristic of a Metal, and a Theory for the Mechanism of Electrode Potential, 494.

MacDonald (J. K. L.) Theory of Uncoupling and Formulæ for the Stark Effect in H_2 , 183.

- McDougall (J.) The Calculation of the Terms of the Optical Spectrum of an Atom with one Series Electron, 550.
- McDowell (C. M.) and Usher (F. L.) The Distribution of Suspended Particles under Gravity, 133.
- McKendrick (A. G.) See Kermack and McKendrick.
- McLennan (J. C.), Burton (A. C.), Pitt (A.) and Wilhelm (J. O.) Further Experiments on Superconductivity with Alternating Currents of High Frequency, 245.
- Massey (H. S. W.) The Passage of Neutrons through Matter, 460.
- Melville (H. W.) The Photochemistry of Phosphine, 374.
- Metals, theory (Wilson), 594.
- Mohamed (A. F.) See Bangham and others.
- Mohr (C. B. O.) and Nicoll (F. H.) Inelastic Electron Scattering in Gases, I, 229.
- Mohr (C. B. O.) and Nicoll (F. H.) The Large Angle Scattering of Electrons in Gases, II, 469.
- Mott (N. F.) See Taylor and Mott.
- Müller (A.) An X-Ray Investigation of Normal Paraffins near their Melting Points, 514.
- Neutrons, passage through matter (Massey), 460.
- Nicoll (F. H.) See Mohr and Nicoll.
- Nitrogen, heat of dissociation (Dutta), 84.
- Nitrous oxide absorption spectrum (Dutta), 84.
- Nottage (M.) See Hardy and Nottage.
- Paraffins, X-ray investigation of normal (Müller), 514.
- Particles suspended under gravity, distribution (McDowell and Usher), 133.
- Phosphine, photochemistry (Melville), 374.
- Pitt (A.) See McLennan and others.
- Plastic flow, mechanism (Andrade and Chalmers), 348.
- Plastic strength of soft materials (Schofield and Blair), 707.
- Plasticity and creep in solids (Jeffreys), 283.
- Potential, surface, measurements (Adam and Harding), 411.
- Potentials, surface of unimolecular films (Schulman and Hughes), 430.
- Powell (C. F.) and Brata (L.) Mobility of Alkali Ions in Gases, 117.
- Propellants, rate of burning (Hunt and Hinds), 696.
- Quartz Piezo-electric plates, vibration revealed by interferometer (Dye), 1.
- Quantum theory, physical principles (Temple), 479.
- Radium C, internal conversion coefficient (Hulme), 643.
- Rao (K. R.) and Badami (J. S.) The Resonance Spectrum of Hydrogen, 540.
- Resistivity of polycrystalline wires (Andrade and Chalmers), 348.
- Ritchie (A.) and Ludlam (E. B.) The Oxidation of Sulphur at Low Pressures, 635.
- Sandeman (I.) Bands due to the Hydrogen Molecule: The $2p^3$ II Bands of Hydrogen, 395.
- Schofield (R. K.) and Blair (G. W. Scott) The Relationship between Viscosity, Elasticity, and Plastic Strength of Soft Materials as Illustrated by some Mechanical Properties of Flour Doughs, 707.
- Schulman (J. H.) and Hughes (A. H.) On the Surface Potentials of Unimolecular Films, IV, 430.

- Shafting, torsion and flexure (Shepherd), 607.
- Shaw (P. E.) and Leavey (E. W. L.) Triboelectricity and Friction, VII, 502.
- Shepherd (W. M.) The Torsion and Flexure of Shafting with Keyways or Cracks, 607
- Smith (J. W.) Dipole Moments and Molecular Structure, III, 154.
- Spectrum, absorption, of nitrous-oxide (Dutta), 84.
- Spectrum, investigations in infra-red region (Cassie and Bailey), 531.
- Spectrum of hydrogen (Rao and Badami), 540.
- Spectrum of krypton, interferometric measurements (Jackson), 147.
- Spectrum, optical, of an atom, calculation of terms (McDougall), 550.
- Spooner (E. C. R.) See McAulay and Spooner.
- Stark effect in H_2 , formulæ (MacDonald), 183.
- Sulphur, oxidation at low pressures (Ritchie and Ludlam), 635.
- Sulphur oxychlorides (Smith), 154.
- Superconductivity with alternating currents of high frequency (McLennan and others), 245.
- Taylor (G. I.) The Viscosity of a Fluid Containing Small Drops of Another Fluid, 41.
- Taylor (H. M.) and Mott (N. F.) A Theory of the Internal Conversion of γ -Rays, 665.
- Taylor (W.) and Weale (A.) The Mechanism of the Initiation and Propagation of Detonation in Solid Explosives, 92.
- Temple (G.) The Physical Principles of the Quantum Theory, 479.
- Thompson (H. W.) See Hadmap and others.
- Thorium B and C, gamma-rays (Ellis), 318.
- Thunderclouds, polarity (Halliday), 205.
- Triboelectricity and friction (Shaw and Leavey), 502.
- Usher (F. L.) See McDowell and Usher.
- Uncoupling, rotational, with application to singlet hydrogen bands (Davidson), 580.
- Uncoupling, Theory (MacDonald), 183.
- Viscosity of a fluid (Taylor), 41.
- Viscosity of soft materials (Schofield and Blair), 707.
- Wave resistance, theory (Havelock), 339.
- Weale (A.) See Taylor and Weale.
- Wilhelm (J. O.) See McLennan and others.
- Wilson (A. H.) The Theory of Metals, I, 594.
- X-ray investigation of normal paraffins (Müller), 514.

I. A. R. L. 75

IMPERIAL AGRICULTURAL RESEARCH
INSTITUTE LIBRARY
NEW DELHI.

[illegible]